

2018

## Exploring Forest Diversity and Ecosystem Services Using Big Data and Empirical Dynamic Modeling

James V Watson III

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Watson, James V III, "Exploring Forest Diversity and Ecosystem Services Using Big Data and Empirical Dynamic Modeling" (2018). *Graduate Theses, Dissertations, and Problem Reports*. 7297.  
<https://researchrepository.wvu.edu/etd/7297>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

# Exploring Forest Diversity and Ecosystem Services Using Big Data and Empirical Dynamic Modeling

James V. Watson, III

Dissertation submitted  
to the Davis College of Agriculture, Natural Resources and Design  
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in  
Forest Resources Science

Jingjing Liang, Ph.D., Chair  
Mo Zhou, Ph.D.  
George Merovich, Ph.D.  
Eungul Lee, Ph.D.  
Jonathan Cumming, Ph.D.

Division of Forestry and Natural Resources

Morgantown, West Virginia  
2018

Keywords: tree species diversity, tree size diversity, forest ecosystem, biodiversity, biological conservation, species richness, productivity, stocking charts, empirical dynamic modeling, convergent cross mapping, ecosystem functioning, climatic conditions

Copyright 2018 James V. Watson, III

## ABSTRACT

### Exploring Forest Diversity and Ecosystem Services Using Big Data and Empirical Dynamic Modeling

James V. Watson, III

Forest ecosystems worldwide harbor the majority of terrestrial biodiversity, and interact intensely with freshwater systems. Forested ecosystems around the globe are experiencing loss in biodiversity. Biodiversity and ecosystem functioning relationship studies have primarily focused on terrestrial systems, and forest biodiversity effects on freshwater systems have been studied to a lesser extent. In addition, successful integration of biological conservation in forest management remains rare and practical guidelines are lacking. Environmental studies often deal with dynamical systems exhibiting changes in environmental factors and their relationships over time. Most of these studies assume that such systems are linear and employ classic time-series approaches for modeling factors and detecting the direction and strength of the relationships. In chapter 1, I analyzed the effects of tree species richness and tree size diversity on the quality of streams and wetlands, based on a joint watershed-forest database that we derived from ground measurements from 1,044 forest plots and 894 stream sites, and a non-parametric random forest model with diversity metrics and other climatic and physiographic factors as explanatory variables. We found a consistent monotonic to positive effect of tree size diversity on watershed quality, but the effect of tree species diversity on watershed quality was less positive. In chapter 2, I addressed the heretofore-ignored implications of the productivity-biodiversity relationship in natural resource management and developed a stand biodiversity capacity (*SBC*) framework as a novel and practical tool set to facilitate the integration of biological conservation in forestry practices. *SBC* was defined as a spatially explicit index to represent the amount of tree species that a forest stand is capable of sustaining. We also developed *SBC* stocking charts for determining the optimal tree species diversity on a local forest stand, given its forest type, basal area, and site productivity. In my last chapter, chapter 3, I contrasted a few classic linear methods which include uni- and multi-variate modeling and Granger Causality test, to empirical dynamic modeling (EDM) methods which include simplex models, s-mapping, and Convergent Cross Mapping (CCM). Results show that EDM provides more predictive power over linear methods with all data lengths, temporal frequencies, and number of variables. Additionally, EDM allows for the analysis of interactive strength of variables across the state space, which could help inform and refine process-based hydrologic models.

## Dedication

I dedicate this dissertation to my mother, Mary Beth Watson, not only for her help and support for the duration of my graduate program, but also for the tremendous amount of time and effort that she put forth through the process of homeschooling me K-12.

I also dedicate this dissertation to my father, James Watson, Jr., and my sisters Abigail Watson and Rebekah Stevenson for their assistance and support throughout my education.

## Acknowledgments

I acknowledge my friends Aaron Carpenter, Aaron Holley, Adam Maltempie, Anthony Pappas, Cory Trego, Darren Wood, Philip Crim, and Wu Ma for their tremendous support throughout my graduate program.

I would also like to acknowledge my Ph.D. program committee members Dr. Eungul Lee, Dr. George Merovich, Dr. Jingjing Liang, Dr. Jonathan Cumming, and Dr. Mo Zhou for their patience, support, and advice as they assisted me throughout my Ph.D. program.

Lastly, I would like to thank my aunt, Esther Morales for her support during the completion of my graduate program at WVU.

## Table of Contents

### Table of Contents

ABSTRACT.....	ii
Dedication .....	iii
Acknowledgments.....	iv
Table of Contents.....	v
Chapter 1: Effects of Biodiversity on Watershed Quality in West Virginia.....	1
Chapter 2: Integrating Biological Conservation in Forest Management with Stand-Biodiversity-Capacity Framework.....	22
Chapter 3: Evaluation of Empirical Dynamic Modeling in a Hydrologic System in West Virginia .....	42

# Chapter 1:

## Effects of Biodiversity on Watershed Quality in West Virginia

### Abstract

The past two decades saw an influx of studies on the biodiversity-ecosystem functioning relationship (BEFR), with overwhelming evidence for a positive effect of biodiversity on ecosystem productivity and services. Forest ecosystems worldwide harbor the majority of terrestrial biodiversity, and interact intensely with freshwater systems. However, BEFR studies have primarily focused on terrestrial systems, and biodiversity effects on freshwater systems remain largely unknown. This study determined the effects of tree species richness and tree size diversity on the quality of streams and wetlands, based on a joint watershed-forest database that we derived from ground measurements from 1,044 forest plots and 894 stream sites, and a non-parametric random forest model with diversity metrics and other climatic and physiographic factors as explanatory variables. We found a consistent monotonic to positive effect of tree size diversity on watershed quality, but the effect of tree species diversity on watershed quality was less negative.

## Introduction

The past two decades saw an increased discussion on the biodiversity-ecosystem functioning relationship (BEFR), that is the effects of biodiversity on the processes and functioning of ecosystems (Cardinale et al. 2012; Naeem et al. 2012). Changes in ecosystem functioning would have both direct and indirect impacts on the products and services that a given ecosystem is capable of providing (Balvanera et al. 2014; Cadotte et al. 2011; Isbell et al. 2015; Mace et al. 2012). Most biodiversity studies completed thus far have shown that biodiversity has a positive impact on ecosystem productivity (e.g. Cardinale et al. 2012; Naeem et al. 2012). However, most of these existing studies have been focused on grassland ecosystems (Cardinale et al. 2012; Loreau et al. 2001). Forest systems are more complex in vertical structures and it can take decades or longer for a forest to complete its successional stages (Barnes et al. 1998), thus making it difficult to relate the grassland findings to forested ecosystems (Symstad et al. 2003).

BEFR studies in forest ecosystems, although fewer in number than grassland studies, show a globally consistent positive effect of biodiversity on forest productivity and ecosystem services (Liang et al. 2016; Paquette and Messier 2011; Zhang et al. 2012). The majority of terrestrial biodiversity is found in the world's forests (FAO 2015) and an intense interaction of freshwater systems and biodiversity is a feature of forest ecosystems (Welsch et al. 2000). The quality of the watersheds within a forested ecosystem is an important component of forest ecosystem functioning. A healthy forest ecosystem typically provides such key processes and services as cleansing water, hydrologic flux and storage, and maintaining of hydrological cycles (Christensen et al. 1996). Plants that are in close proximity to waterbodies improve water temperature (Karr and Schlosser 1978; Lowrance et al. 1985) and help to stabilize banks surrounding the waterbody (Karr and Schlosser 1978; Lowrance et al. 1985). Trees also impact



hydrological processes at all stages of forest development (Tabacchi et al. 2000), reduce stream sedimentation (Karr and Schlosser 1978; Lowrance et al. 1985; Lowrance et al. 1984), and control the movement of nutrients into bodies of water (Karr and Schlosser 1978; Lowrance et al. 1985; Lowrance et al. 1984; Peterjohn and Correll 1984; Tabacchi et al. 2000). In summary, stream biotic and chemical quality is strongly influenced by nearby biodiversity due to the effects of vegetation on aquatic food webs, water temperature, nutrient cycle and decomposition, and instream sedimentation (Karr and Schlosser 1978).

Studies of biodiversity and ecosystem functioning have primarily focused on terrestrial systems and to a lesser extent, and biodiversity effects on freshwater systems remain largely unknown (Hooper et al. 2005). Out of 5.2 km of waterways in the 48 contiguous states in the United States, 2% are of high enough quality to receive a status of protection from the U.S. government (Benke 1990). Thus, it is important to determine the effects of terrestrial biodiversity on freshwater systems and how diversity of trees can be managed to promote watershed quality. The primary objective of this study was to determine the effects of tree species richness and tree size diversity on the quality of streams and wetlands, based on a joint watershed-forest database that we created for northern West Virginia.

#### Data and Methods

Three types of data were used in this study— forest data, climate data, and water quality data. Forest data was obtained from the Forest Inventory and Analysis (FIA) dataset of the United States of America (Woudenberg et al. 2011) and the Global Land Cover Facility (Sexton et al. 2013). Climate data was obtained from a climate dataset developed by Hijmans et. al. (2005). All water quality data was provided by the West Virginia Department of Environmental

Protection (WV DEP) Division of Water and Waste Management Watershed Assessment Program (WVDEP 2012).

With the exception of percent tree cover, forest data were observed data collected from field surveys on permanent FIA sample plots located across the United States (Woudenberg et al. 2011). Under the current FIA sampling regime, a portion of FIA plots of each state are re-inventoried each year (Gillespie 1999). On each FIA plot, individual tree level attributes such as diameter-at-breast-height (DBH) and species were recorded, which we derived attributes of tree species diversity, species composition, and tree size diversity. Tree species diversity was the total number of woody plant species recorded on an FIA plot, and tree size diversity was the coefficient of variation of DBH values within FIA plots. In this study, species composition consists of seven species groups. These groups were created based on species' taxonomic features and are as follows: white oak species (*Quercus* – *Quercus*, QQ), red oak species (*Quercus* – *Lobatae*, QL), *Juglandaceae* (JD), *Sapindaceae* (SD), gymnosperms (GS), *Fagus* (FG), and other angiosperms (OA). Grouping species together not only allowed for simplicity of the study, but was necessary due to computational constraints (Ma et al. 2016).

Tree cover percentage is defined as the estimate of the percentage of ground in a 30 x 30 meter pixel that is covered by woody plants greater than 5 meters in height. After obtaining the tree cover percentage from the Global Land Cover Facility, the tree cover percentage data was overlain in ArcGis 10.3 by 24000 scale reachsheds. The tree cover percentage of each reachshed containing watershed sampling data was calculated via the zonal statistics feature in ArcGIS 10.3 and appended to the water sampling data. The result was that each water sampling site data contained the percentage of tree cover for the reachshed containing the water sampling site.

Historical climate data came from interpolated climate surfaces for global land areas with the exception of Antarctica (Hijmans et al. 2005). The climate surfaces were overlaid against the FIA plot coordinates to obtain the historical mean annual temperature and total annual precipitation of the FIA locations.

Water quality data were collected by the Watershed Assessment Branch (WAB) of WV DEP, with a primary goal to evaluate watershed quality in the State of West Virginia in accordance with the Federal Clean Water Act (CWA)(WVDEP 2012). The WAB measures all the 32 eight digit hydrologic unit watersheds of the state on a five-year cycle (Fig. 1). Thus, around 20 percent of these watersheds are sampled each year. The WAB sampling entails the collection of biological and water samples, as well as an assessment of instream and riparian habitat (WVDEP 2012).

We joined the watershed and forest data, based on the fluvial distance between forest and watershed sampling sites (Fig 2). Using the network analyst feature of ArcGis 10.3 (ESRI 2014), we calculated the fluvial distance between an FIA and a WAB sampling sites, that is, the distance along the stream network to the WAB sampling site from an FIA plot (Fig. 2). It was found that tree species diversity was spatially auto correlated with other variables of our dataset at a distance of 10 km. Only data from watershed sampling sites located downstream of a particular FIA plot was combined with data from said FIA plot. Data from a total of 1,044 FIA plots, which were connected via the stream network to watershed sampling sites, was combined with water quality data (Fig. 1). This data joining method ensures that FIA data were a good representation of the forest conditions with regard to a particular watershed (Fig. 2).

Our joint forest-watershed database, derived from direct measurements on 1,044 FIA plots and 894 WAB sites, consists 6,719 matching records of watershed quality and associated

forest and climatic conditions. We selected 23 attributes for the analysis of forest-water quality relationship (Table 1). All the explanatory variables were derived from the FIA dataset, the tree cover dataset of the Global Land Cover Facility, and the historical climate data set. The response variables representing metrics of water quality, Rapid Visual Habitat Assessment (RVHA), water chemistry indices (PC1, PC2), and West Virginia Stream Condition Index (WVSCI) were obtained directly (RVHA and WVSCI) and derived (PC1 and PC2) from the WAB dataset (Table 1).

RVHA is a score reflecting the quality of key instream and riparian habitat components, such as the epifaunal substrate, embeddedness, velocity, channel alteration, sediment deposition, frequency of riffles, channel flow status, bank stability, bank vegetative protection, and riparian vegetative zone width (Barbour et al. 1999). RVHA scores assigned to each of the key components range from 0 to 20, and the total score from 10 components (ranged from 0 to 200, with 200 being the best quality) can then be compared (Barbour et al. 1999).

WVSCI is derived from a bio-assessment procedure in which the biological condition of a stream is compared to a reference condition (Green and Swietlik 2000). The reference condition is an aggregate of conditions found in streams of the best biological conditions in a region. WVSCI is comprised of six metrics (Green and Swietlik 2000) representative of the function and composition of the macroinvertebrate assemblages. The metrics included in the WVSCI are EPT taxa, total taxa, percent of EPT, percent of Chironomidae, percent of top two dominant taxa, and HBI (family biotic index). Each of these six metrics are assigned a score between 0 (poorest quality) and 100 (highest quality). The scores are then standardized and the average of the six final scores of the metrics was assigned as the final WVSCI score (Green and Swietlik 2000).

The water chemistry indices are principal components one and two from completing a principal component analysis (PCA) on water chemistry variables. The variables used in the creation of the water chemistry indices are pH, dissolved oxygen, conductivity, fecal content, alkalinity, sulfate, chloride, total suspended solids, total phosphorus, dissolved aluminum, total calcium, dissolved copper, dissolved iron, dissolved zinc, and nitrate.

We used random forest (Breiman 2001) and linear stepwise regression to determine the effects of tree species and size diversity on watershed quality. The non-parametric random forest approach is particularly useful to our study, because our watershed and forest data have an unknown distribution and random forest runs efficiently on large databases, such as our joint watershed-forest database. The random forest approach uses bootstrapping of the sample data to build decision trees and then randomly chooses a subset of explanatory variables when splitting at nodes (Genuer et al. 2010). Thus, the Random Forest method adds an additional component of randomness in comparison to the traditional bagging of classification trees (Liaw and Wiener 2002).

Tree size diversity and tree species diversity were held constant at their sample means for both linear regression and random forest analysis. To achieve this for the random forest model, we created partial dependency plots, a technique useful to show the effect of tree species diversity and size diversity on the dependent variable, while all other independent variables were held constant at the sample mean. Both the random forest approach and a linear stepwise regression model were completed for the relationship between tree species diversity, tree size diversity and each of the four response variables.

## Results

The random forest model showed a monotonic effect of tree size diversity on watershed quality. Other things being equal, RVHA scores remained close to 147 (Fig. 3 A), WVSCI scores remained around 71 (Fig. 4 A), PC1 scores close to -1 (Fig. 5A), and PC2 scores close to -0.2 (Fig. 6A) as tree size diversity increased from 0.00 to 1.23.

The random forest model also showed a monotonic effect of tree species diversity on watershed quality. Other things being equal, as tree species diversity increased, RVHA scores in general remained close to 147 (Fig. 3 B), WVSCI scores held close to 71 (Fig. 4 B), PC1 scores close to -1 (Fig. 5B), and PC2 scores close to -0.2 (Fig. 6B).

The linear regression model showed a mostly positive effect of tree size diversity on watershed quality. Other things being equal, RVHA scores increased from approximately 138 to around 152 (Fig. 3 C), and WVSCI scores increased from around 56 to approximately 70 (Fig. 4 C), as tree size diversity increased from 0.00 to 1.23. A slight increase in PC1 scores was shown (Fig. 5 C) as tree size diversity increased. PC2 scores decreased from approximately 0.3 to around -0.1 with increasing tree size diversity. A decrease in PC scores does not necessarily indicate a decreased quality of water with increasing tree size diversity.

The linear regression model showed both a positive and negative effect of tree species diversity on watershed quality. Other things being equal, as tree species diversity increased, RVHA scores in general decreased from 144 to 130 (Fig. 3 D). Other things being equal, as tree species diversity increased, WVSCI scores also increased from 65 to 87 (Fig. 4 D). As tree species diversity increased, PC1 scores decreased from around -1 to -2 and PC2 scores increased sharply from approximately -0.1 to 0.03. As mentioned previously, a decrease or increase in PC scores should not be assumed to be correlated with an increase or decrease of water quality.

## Discussion and Conclusion

The consistent positive effect of tree size diversity on watershed quality agreed with knowledge that forest structural diversity is essential in enhancing ecosystem stability (Hansen et al. 1991; Sharitz et al. 1992), quality (Hansen et al. 1991), and productivity (Crow 1989). The amount and size of fallen trees varies with stages of forest succession due to differences in tree size across stages of succession (Hansen et al. 1991). Coarse woody debris (CWD) are important to stream ecology. Aquatic organisms utilize CWD for habitat (Anderson and Sedell 1979) and fallen trees assist with stabilizing stream banks (Keller and Swanson 1979). The quality of habitat provided to aquatic organisms and the stability given to stream banks from CWD would be affected by size of CWD as decomposition rates of CWD decrease with CWD size (Triska and Cromack Jr 1980). In addition, tree size may also affect decomposition rates, as evidenced by a study in the southern Appalachian Mountains, where it was found that forests at earlier stages of succession contained species whose leaves decomposed more quickly compared to the species found in forests at later stages of forest succession (Webster et al. 1983).

Along PC1, higher values are associated with higher water pH, total suspended solids, and greater conductivity levels of the water. Increased values of PC2 are representative of higher levels of aluminum, iron, and acidity. The observation that PC2 scores decreased as diversity of tree size increased may be explained by the possibility that older, more mature forests are generally better at reducing runoff and filtering chemicals from runoff prior to it entering streams. That PC1 scores would decrease with increasing tree species diversity is of notable interest, yielding the idea that increasing tree species diversity may potentially reduce water acidity and conductivity levels.

The positive and negative effects of tree species diversity on watershed quality reflects the fact that changes in plant species diversity may influence ecosystem stability (Loreau et al. 2001; Yachi and Loreau 1999). Tree species diversity has the potential to affect both the decomposition rate of leaf litter in streams and also the species richness of invertebrate populations in the stream (Leroy and Marks 2006). Invertebrates classified by their feeding habits as shredders, require that leaf litter be conditioned before feeding on it. The time of the conditioning process, carried out by microbes, varies by tree species (Cummins et al. 1989). The rate at which leaf litter decays varies among different species of trees (Ostrofsky 1997; Webster and Benfield 1986; Webster et al. 1999). Leaves of different tree species vary in nutrient content and toughness (Melin 1930) and these differences could be connected to the discrimination of certain leaf litter by shredders (Graça 2001). It has been found that invertebrate biomass was higher with increased diversity of tree species litter (Leff and McArthur 1989).

There is need for further analysis on the role of tree species richness and tree size diversity on stream quality. Streams and the overall watersheds that they are a component of are important for many reasons ranging from productivity of the ecosystem as a whole, to providing recreational options to humans. With a threat of both species loss and watershed degradation, it is important that we attempt to fully understand the components of an ecosystem and processes that may affect stream/watershed quality and productivity.



## References:

- Anderson, N.H., and Sedell, J.R. 1979. Detritus processing by macroinvertebrates in stream ecosystems. *Annual review of entomology* **24**(1): 351-377.
- Balvanera, P., Siddique, I., Dee, L., Paquette, A., Isbell, F., Gonzalez, A., Byrnes, J., O'Connor, M.I., Hungate, B.A., and Griffin, J.N. 2014. Linking biodiversity and ecosystem services: current uncertainties and the necessary next steps. *Bioscience* **64**(1): 49-57.
- Barbour, M.T., Gerritsen, J., Snyder, B., and Stribling, J. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers. USEPA, Washington.
- Barnes, B.V., Zak, D.R., Denton, S.R., and Spurr, S.H. 1998. *Forest Ecology*. 4th ed. John Wiley & Sons, Inc., Lexington, KY.
- Benke, A.C. 1990. A perspective on America's vanishing streams. *Journal of the North American Benthological Society* **9**(1): 77-88.
- Breiman, L. 2001. Random forests. *Machine learning* **45**(1): 5-32.
- Cadotte, M.W., Carscadden, K., and Mirotchnick, N. 2011. Beyond species: functional diversity and the maintenance of ecological processes and services. *J. Appl. Ecol.* **48**(5): 1079-1087. doi: 10.1111/j.1365-2664.2011.02048.x.
- Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., and Naeem, S. 2012. Biodiversity loss and its impact on humanity. *Nature* **486**(7401): 59-67.
- Christensen, N.L., Bartuska, A.M., Brown, J.H., Carpenter, S., D'Antonio, C., Francis, R., Franklin, J.F., MacMahon, J.A., Noss, R.F., and Parsons, D.J. 1996. The report of the Ecological Society of America committee on the scientific basis for ecosystem management. *Ecological Applications* **6**(3): 665-691.
- Crow, T.R. 1989. Biological diversity and silvicultural systems. *In* *Proceedings of the National Silviculture Workshop: Silvicultural Challenges*. pp. 180-184.
- Cummins, K.W., Wilzbach, M.A., Gates, D.M., Perry, J.B., and Taliaferro, W.B. 1989. Shredders and riparian vegetation. *BioScience* **39**(1): 24-30.
- ESRI. 2014. ArcGIS Desktop: Release 10.3. Environmental Systems Research Institute, Redlands, California.
- FAO. 2015. *Global Forest Resources Assessment 2015 - How are the world's forests changing?*. Food and Agriculture Organization of the United Nations.
- Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. 2010. Variable selection using random forests. *Pattern Recognition Letters* **31**(14): 2225-2236.
- Gillespie, A.J. 1999. Rationale for a national annual forest inventory program. *Journal of Forestry* **97**(12): 16-20.
- Graça, M.A. 2001. The role of invertebrates on leaf litter decomposition in streams—a review. *International Review of Hydrobiology* **86**(4-5): 383-393.
- Green, J., and Swietlik, W. 2000. A STREAM CONDITION INDEX FOR WEST VIRGINIA WADEABLE STREAMS.
- Hansen, A.J., Spies, T.A., Swanson, F.J., and Ohmann, J.L. 1991. Conserving biodiversity in managed forests. *BioScience* **41**(6): 382-392.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**(15): 1965-1978.
- Hooper, D., Chapin III, F., Ewel, J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J., Lodge, D., Loreau, M., and Naeem, S. 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological monographs* **75**(1): 3-35.
- Isbell, F., Tilman, D., Polasky, S., and Loreau, M. 2015. The biodiversity-dependent ecosystem service debt. *Ecol. Lett.* **18**(2): 119-134. doi: 10.1111/ele.12393.

Karr, J.R., and Schlosser, I.J. 1978. Water Resources and the Land-Water Interface. *science* **201**: 21.

Keller, E.A., and Swanson, F.J. 1979. Effects of large organic material on channel form and fluvial processes. *Earth Surface Processes and Landforms* **4**(4): 361-380.

Leff, L.G., and McArthur, J.V. 1989. The effect of leaf pack composition on processing: a comparison of mixed and single species packs. *Hydrobiologia* **182**(3): 219-224.

Leroy, C.J., and Marks, J.C. 2006. Litter quality, stream characteristics and litter diversity influence decomposition rates and macroinvertebrates. *Freshwater biology* **51**(4): 605-617.

Liang, J., Crowther, T.W., Picard, N., Wiser, S., Zhou, M., Alberti, G., Schulze, E.-D., McGuire, A.D., Bozzato, F., Pretzsch, H., de-Miguel, S., Paquette, A., Hérault, B., Scherer-Lorenzen, M., Barrett, C.B., Glick, H.B., Hengeveld, G.M., Nabuurs, G.-J., Pfautsch, S., Viana, H., Vibrans, A.C., Ammer, C., Schall, P., Verbyla, D., Tchebakova, N., Fischer, M., Watson, J.V., Chen, H.Y.H., Lei, X., Schelhaas, M.-J., Lu, H., Gianelle, D., Parfenova, E.I., Salas, C., Lee, E., Lee, B., Kim, H.S., Bruelheide, H., Coomes, D.A., Piotta, D., Sunderland, T., Schmid, B., Gourlet-Fleury, S., Sonké, B., Tavani, R., Zhu, J., Brandl, S., Vayreda, J., Kitahara, F., Searle, E.B., Neldner, V.J., Ngugi, M.R., Baraloto, C., Frizzera, L., Bałazy, R., Oleksyn, J., Zawila-Niedzwiecki, T., Bouriaud, O., Bussotti, F., Finér, L., Jaroszewicz, B., Jucker, T., Valladares, F., Jagodzinski, A.M., Peri, P.L., Gonmadje, C., Marthy, W., O'Brien, T., Martin, E.H., Marshall, A.R., Rovero, F., Bitariho, R., Niklaus, P.A., Alvarez-Loayza, P., Chamuya, N., Valencia, R., Mortier, F., Wortel, V., Engone-Obiang, N.L., Ferreira, L.V., Odeke, D.E., Vasquez, R.M., Lewis, S.L., and Reich, P.B. 2016. Positive biodiversity-productivity relationship predominant in global forests. *Science* **354**(6309). doi: 10.1126/science.aaf8957.

Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. *R news* **2**(3): 18-22.

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J., Hector, A., Hooper, D., Huston, M., Raffaelli, D., and Schmid, B. 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**(5543): 804-808.

Lowrance, R., Leonard, R., and Sheridan, J. 1985. Managing riparian ecosystems to control nonpoint pollution. *Journal of soil and water conservation* **40**(1): 87-91.

Lowrance, R., Todd, R., Fail, J., Hendrickson, O., Leonard, R., and Asmussen, L. 1984. Riparian forests as nutrient filters in agricultural watersheds. *BioScience* **34**(6): 374-377.

Ma, W., Liang, J., Cumming, J.R., Lee, E., Welsh, A.B., Watson, J.V., and Zhou, M. 2016. Fundamental shifts of central hardwood forests under climate change. *Ecological Modelling* **332**: 28-41.

Mace, G.M., Norris, K., and Fitter, A.H. 2012. Biodiversity and ecosystem services: a multilayered relationship. *Trends Ecol. Evol.* **27**(1): 19-26.

Melin, E. 1930. Biological decomposition of some types of litter from North American forests. *Ecology* **11**(1): 72-101.

Naeem, S., Duffy, J.E., and Zavaleta, E. 2012. The Functions of Biological Diversity in an Age of Extinction. *Science* **336**(6087): 1401-1406. doi: 10.1126/science.1215855.

Ostrofsky, M. 1997. Relationship between chemical characteristics of autumn-shed leaves and aquatic processing rates. *Journal of the North American Benthological Society* **16**(4): 750-759.

Paquette, A., and Messier, C. 2011. The effect of biodiversity on tree productivity: from temperate to boreal forests. *Global Ecology and Biogeography* **20**(1): 170-180.

Peterjohn, W.T., and Correll, D.L. 1984. Nutrient dynamics in an agricultural watershed: observations on the role of a riparian forest. *Ecology* **65**(5): 1466-1475.

Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., and DiMiceli, C. 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth* **6**(5): 427-448.

Sharitz, R.R., Boring, L.R., Lear, D.H.V., and Pinder, J.E. 1992. Integrating ecological concepts with natural resource management of southern forests. *Ecological Applications* **2**(3): 226-237.

Symstad, A.J., Chapin, F.S., Wall, D.H., Gross, K.L., Huenneke, L.F., Mittelbach, G.G., Peters, D.P., and Tilman, D. 2003. Long-term and large-scale perspectives on the relationship between biodiversity and ecosystem functioning. *Bioscience* **53**(1): 89-98.

Tabacchi, E., Lambs, L., Guilloy, H., Planty-Tabacchi, A.M., Muller, E., and Decamps, H. 2000. Impacts of riparian vegetation on hydrological processes. *Hydrological processes* **14**(16-17): 2959-2976.

Triska, F.J., and Cromack Jr, K. 1980. The role of wood debris in forests and streams. *Forests: fresh perspectives from ecosystem analysis*. Oregon State University Press, Corvallis, Oregon, USA: 171-190.

Webster, J., and Benfield, E. 1986. Vascular plant breakdown in freshwater ecosystems. *Annual review of ecology and systematics* **17**(1): 567-594.

Webster, J., Benfield, E., Ehrman, T., Schaeffer, M., Tank, J., Hutchens, J., and D'angelo, D. 1999. What happens to allochthonous material that falls into streams? A synthesis of new and published information from Coweeta. *Freshwater biology* **41**(4): 687-705.

Webster, J., Gurtz, M., Hains, J., Meyer, J., Swank, W., Waide, J., and Wallace, J. 1983. Stability of stream ecosystems. *Stream ecology*: 355-395.

Welsch, D.J., Hornbeck, J.W., Verry, E.S., Dolloff, A., and Greis, J.G. 2000. Riparian area management: themes and recommendations.

Woudenberg, S., Conkling, B., O'Connell, B., LaPoint, E., Turner, J., Waddell, K., Boyer, D., Christensen, G., and Ridley, T. 2011. The Forest Inventory and Analysis Database: description and users manual version 5.1 for Phase 2. US Department of Agriculture, Forest Service.

WVDEP. 2012. (West Virginia Department of Environmental Protection). Watershed Branch 2012 Standard Operating Procedures. Division of Water and Waste Management, Watershed Branch, Charleston, WV.

Yachi, S., and Loreau, M. 1999. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proceedings of the National Academy of Sciences* **96**(4): 1463-1468.

Zhang, Y., Chen, H.Y.H., and Reich, P.B. 2012. Forest productivity increases with evenness, species richness and trait variation: a global meta-analysis. *J. Ecol.* **100**(3): 742-749. doi: 10.1111/j.1365-2745.2011.01944.x.

Table 1. Units and definitions of WV DEP and FIA attributes.

Units		Definition
<b>Response Variable</b>		
<i>H</i>		<i>Rapid Visual Habitat Assessment</i>
<i>W</i>		<i>West Virginia Stream Condition Index</i>
<i>PC1</i>		<i>Principal Component one</i>
<i>PC2</i>		<i>Principal Component two</i>
<b>Explanatory Variable</b>		
<i>E</i>	10 <sup>3</sup> m	FIA plot elevation
<i>AGE</i>	year	Stand age
<i>BA</i>	m <sup>2</sup> ha <sup>-1</sup>	Stand basal area
<i>DCV</i>		Tree size diversity (in terms of coefficient of variation)
<i>SP</i>		Tree species diversity (in terms of number of species on plot)
<i>C<sub>1</sub></i>	10 <sup>6</sup> g	Carbon of litter and organic matter
<i>C<sub>2</sub></i>	10 <sup>6</sup> g	Carbon of dead trees
<i>C<sub>3</sub></i>	10 <sup>6</sup> g	Carbon of the understory
<i>TC</i>	%	Tree cover (as a percentage of the reachshed covered)
<i>T</i>	°C	Mean Annual Temperature
<i>P</i>	mm	Total Annual Precipitation
<i>L</i>	km	Fluvial distance between FIA plot and watershed sampling site
<i>QQ</i>	m <sup>2</sup> ha <sup>-1</sup>	Basal area of white oak species ( <i>Quercus-Quercus</i> )
<i>QL</i>	m <sup>2</sup> ha <sup>-1</sup>	Basal area of red oak species ( <i>Quercus-Lobatae</i> )
<i>JD</i>	m <sup>2</sup> ha <sup>-1</sup>	Basal area of <i>Juglandaceae</i>
<i>SD</i>	m <sup>2</sup> ha <sup>-1</sup>	Basal area of <i>Sapindaceae</i>
<i>GS</i>	m <sup>2</sup> ha <sup>-1</sup>	Basal area of <i>Gymnosperms</i>
<i>FG</i>	m <sup>2</sup> ha <sup>-1</sup>	Basal area of <i>Fagus</i>
<i>OA</i>	m <sup>2</sup> ha <sup>-1</sup>	Basal area of other <i>Angiosperms</i>

Table 2. Summary statistics of variables used in this study, based on 11,554 observations. Std: Standard Deviation.

Variable	Mean	Std.	Max.	Min.	<i>n</i>
<i>E</i>	0.74	0.29	1.41	0.29	6,719
<i>AGE</i>	56.14	32.89	140.00	0.00	6,719
<i>BA</i>	24.60	11.00	55.51	0.09	6,719
<i>DCV</i>	0.51	0.15	1.24	0.00	6,719
<i>SP</i>	7.68	2.64	17.00	1.00	6,719
<i>C<sub>1</sub></i>	78.29	19.00	140.89	57.18	6,719
<i>C<sub>2</sub></i>	13.36	4.87	48.78	1.08	6,719
<i>C<sub>3</sub></i>	1.78	0.15	3.92	1.26	6,719
<i>TC</i>	48.67	14.28	74.98	10.88	6,719
<i>T</i>	9.21	1.07	11.00	7.10	6,719
<i>P</i>	108.88	8.94	129.00	89.00	6,719
<i>L</i>	52.44	47.26	220.73	0.01	6,719
<i>QQ</i>	2.22	5.05	31.00	0.00	6,719
<i>QL</i>	2.32	4.93	39.83	0.00	6,719
<i>JD</i>	1.16	2.66	24.61	0.00	6,719
<i>SD</i>	8.53	7.14	48.00	0.00	6,719
<i>GS</i>	0.79	2.76	41.00	0.00	6,719
<i>FG</i>	1.87	3.83	24.00	0.00	6,719
<i>OA</i>	12.24	9.95	54.73	0.00	6,719
<i>H</i>	143.00	25.67	188.00	0.00	6,719
<i>W</i>	65.60	19.74	99.61	10.63	6,719
<i>PC1</i>	-0.66	1.57	4.02	-4.99	2,777
<i>PC2</i>	-0.03	0.51	1.23	-4.54	2,777

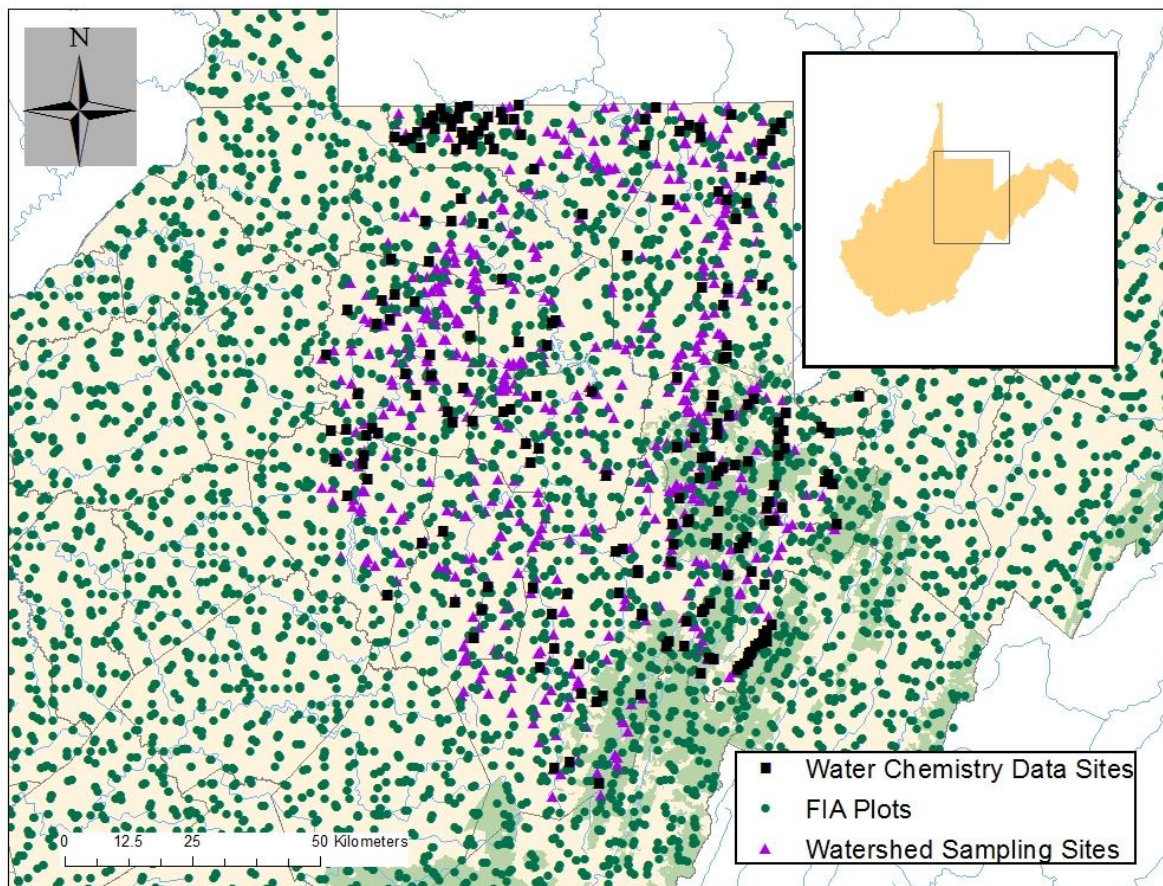


Figure 1. Location of Forest Inventory and Analysis (FIA) plots and watershed sampling sites across the study area in northern West Virginia.



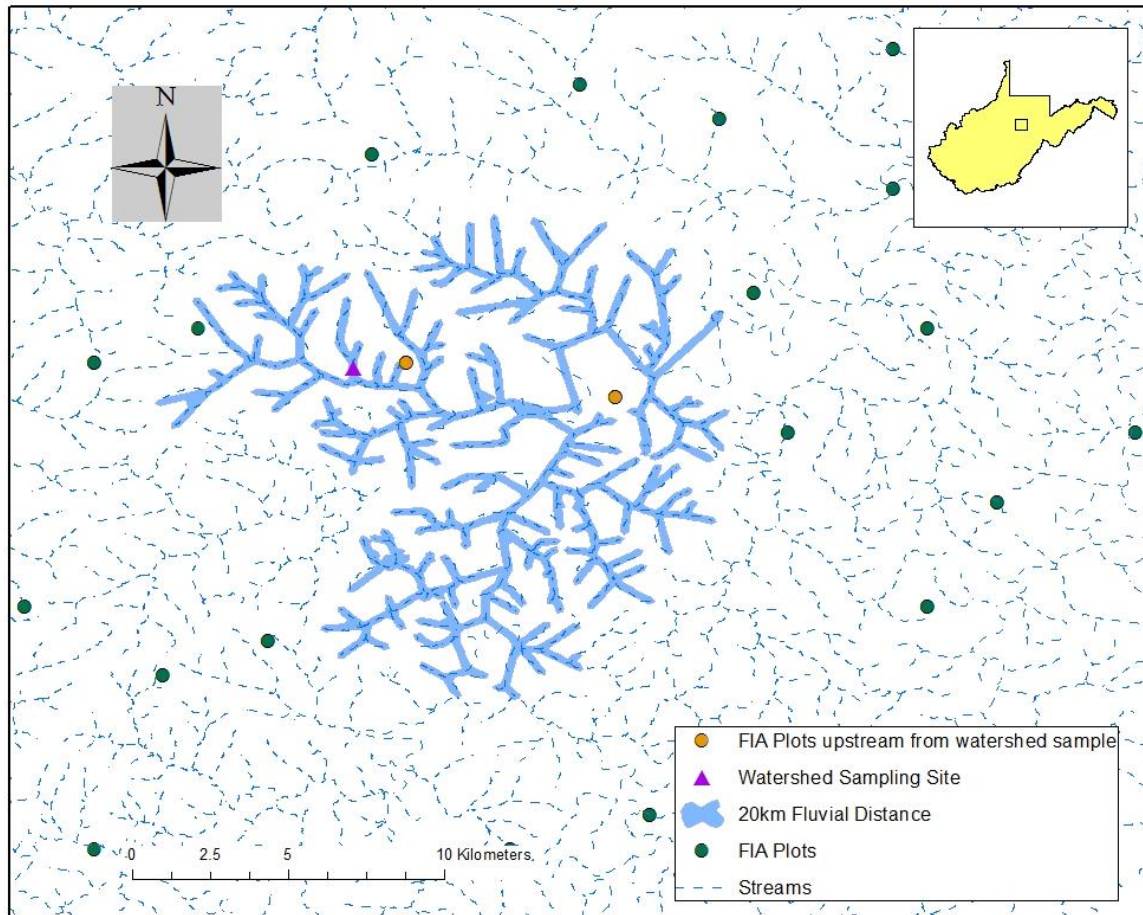


Figure 2. Fluvial distance between Forest Inventory and Analysis (FIA) plots and a watershed sampling site. The fluvial distance is shown by colored lines which follow the stream network between the FIA plots and watershed sampling sites.

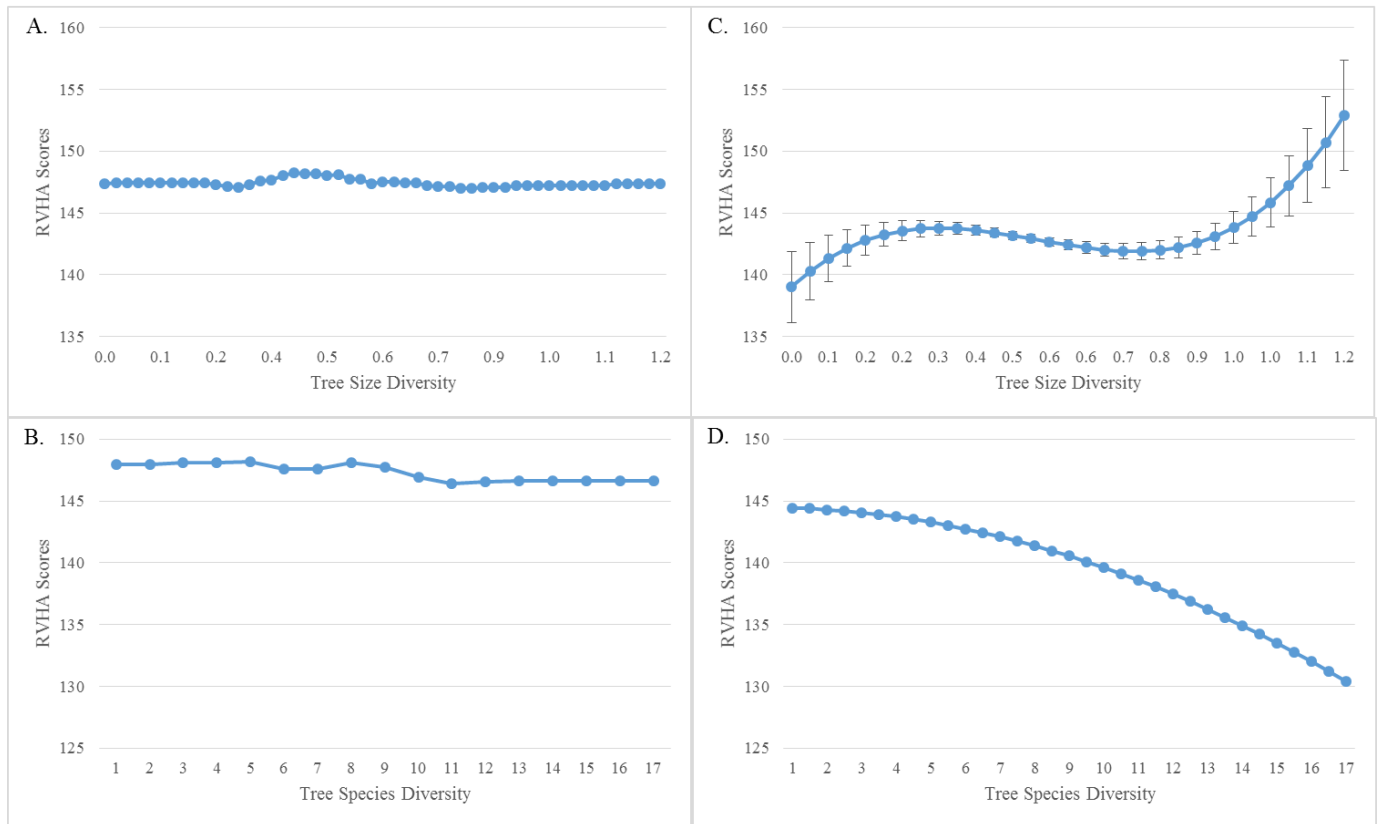


Figure 3. Effects of Tree Size Diversity (A.) and Tree Species Diversity (B.) on RVHA Scores utilizing Random Forest. Effects of Tree Size Diversity (C.) and Tree Species Diversity (D.) on RVHA Scores using Linear Regression.



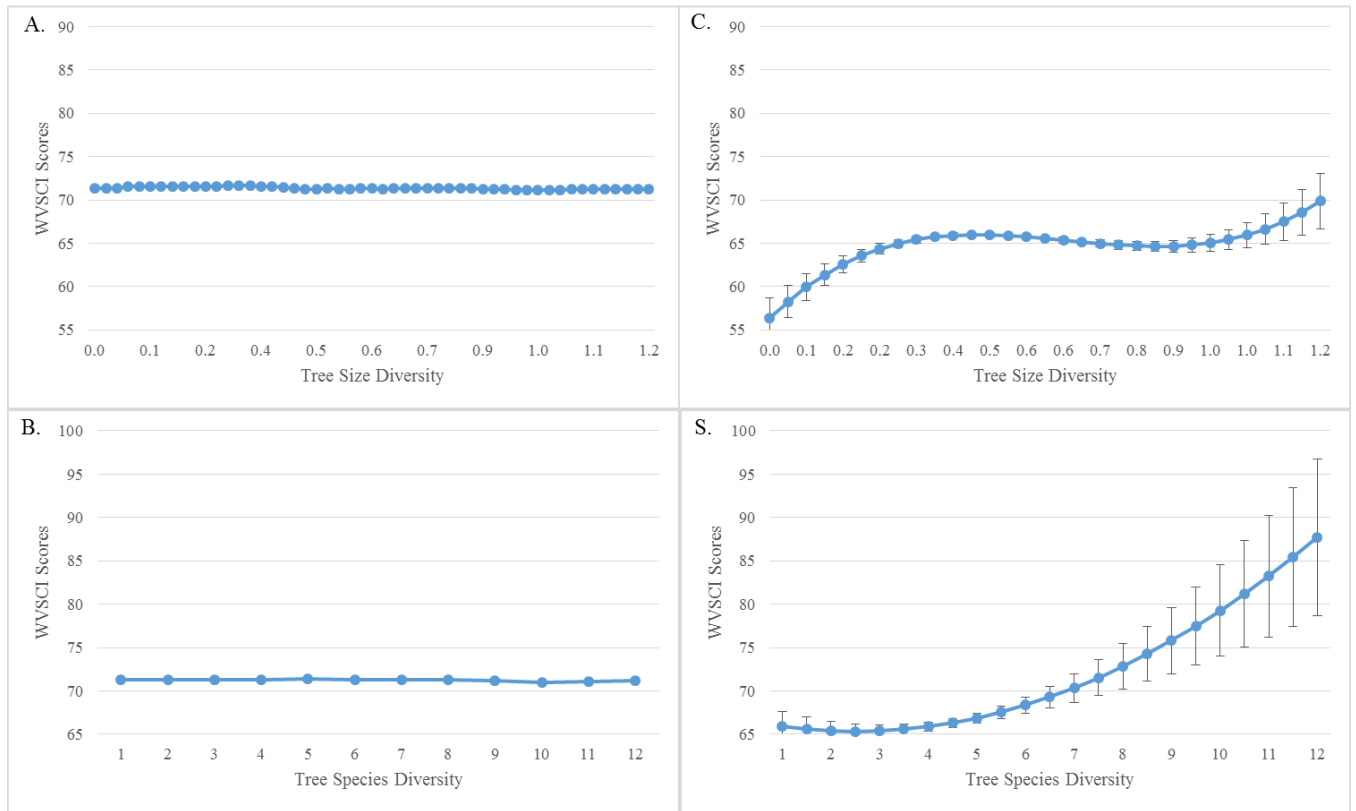


Figure 4. Effects of Tree Size Diversity (A.) and Tree Species Diversity (B.) on WVSCI Scores utilizing Random Forest. Effects of Tree Size Diversity (C.) and Tree Species Diversity (D.) on WVSCI Scores using Linear Regression.

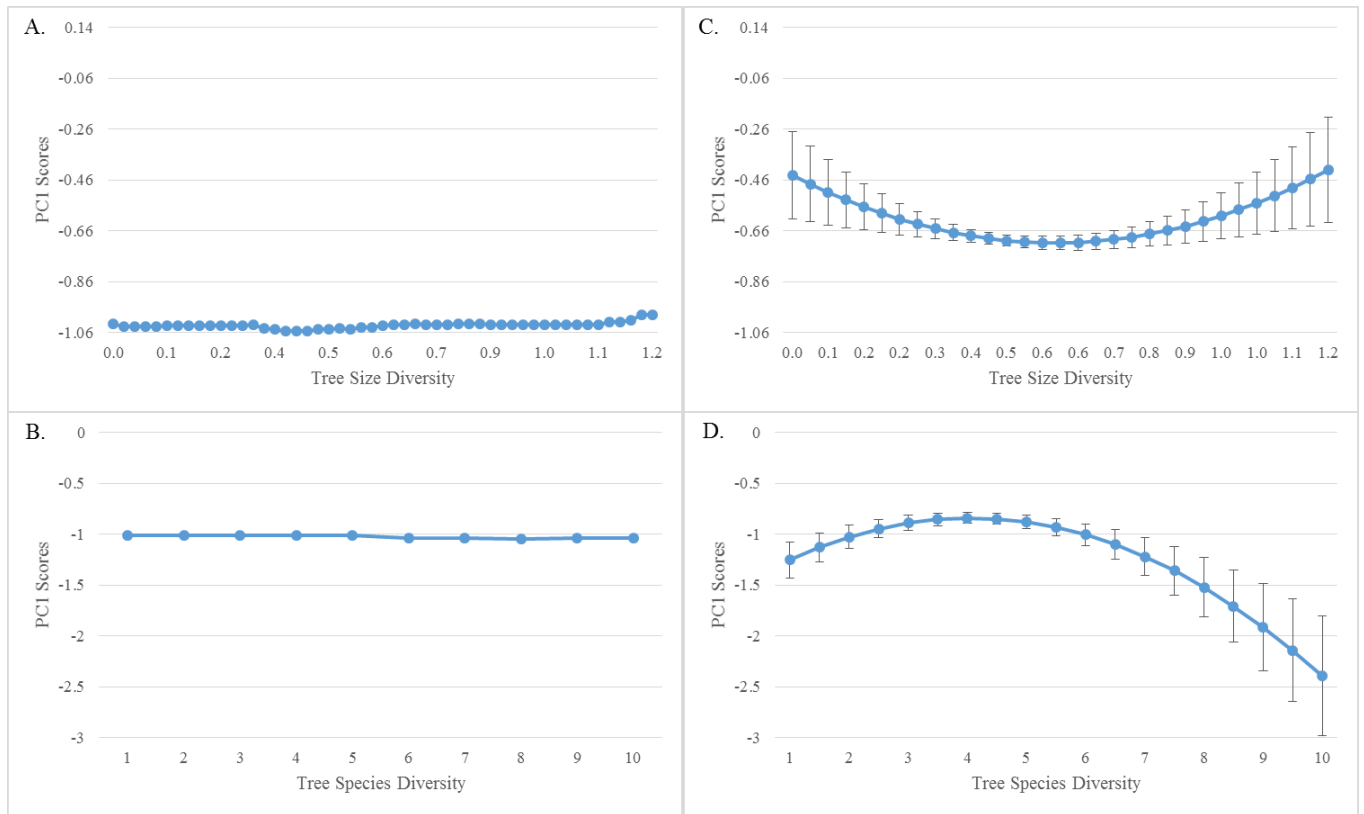


Figure 5. Effects of Tree Size Diversity (A.) and Tree Species Diversity (B.) on PC1 Scores utilizing Random Forest. Effects of Tree Size Diversity (C.) and Tree Species Diversity (D.) on PC1 Scores using Linear Regression.

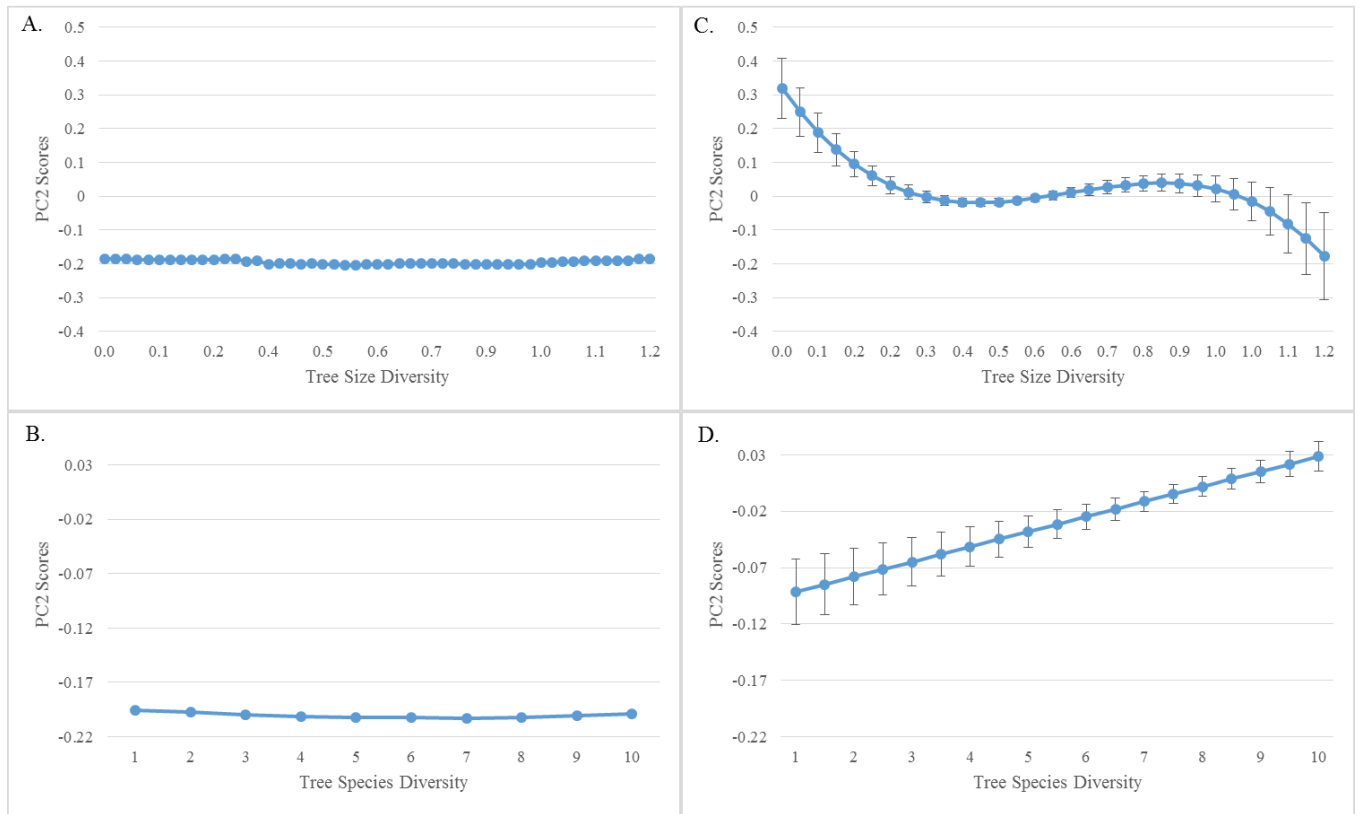


Figure 6. Effects of Tree Size Diversity (A.) and Tree Species Diversity (B.) on PC2 Scores utilizing Random Forest. Effects of Tree Size Diversity (C.) and Tree Species Diversity (D.) on PC2 Scores using Linear Regression.

## Chapter 2:

### Integrating Biological Conservation in Forest Management with Stand-Biodiversity-Capacity Framework

#### **Abstract**

Forested ecosystems around the globe are experiencing loss in biodiversity. Meanwhile, successful integration of biological conservation in forest management remains rare and practical guidelines are lacking. In this paper, we addressed the heretofore-ignored implications of the productivity-biodiversity relationship in natural resource management and developed a stand biodiversity capacity (*SBC*) framework as a novel and practical tool set to facilitate the integration of biological conservation in forestry practices. *SBC* was defined as a spatially explicit index to represent the amount of tree species that a forest stand is capable of sustaining. To demonstrate the generality of the *SBC* framework, we applied this approach to 16 forest types in the United States and three additional forest types in China. Digital maps were created for the contiguous U.S. states, southern Alaska, and the Wangqing Forestry Bureau of China, displaying current biodiversity stocking and *SBC*. We also developed *SBC* stocking charts for determining the optimal tree species diversity on a local forest stand, given its forest type, basal area, and site productivity. These *SBC* tools would be useful in integrating conservation efforts into forest management practices to various forest types on a broader scale.

#### *Key words:*

Biodiversity, forest ecosystem, biological conservation, forestry, species richness, forest productivity, stocking charts

## INTRODUCTION

The world's forested ecosystems house the majority of the Earth's terrestrial biodiversity (FAO 2015). Approximately 80 percent of the plant biomass on the planet is stored in forested ecosystems (Pan et al. 2013), and forests contribute 75 percent of the total terrestrial primary productivity on the globe (Pan et al. 2013). Forests provide many critical ecosystem services for humans and are crucial for socio-economic development (FAO 2011, Liang et al. 2016a). As biodiversity continues to decline in the world's forests (Butchart et al. 2010), integrating conservation with forest resource management is becoming increasingly important. Nevertheless, existing methods of intergrading conservation into forest management are largely unsuccessful due to a lack of practical guidelines (Pukkala 2002, Lindenmayer et al. 2006). Thus, practical forest resource management tools to assist with biodiversity conservation are in dire need.

Understanding the role of productivity in regulating plant diversity is critical to the integration of biodiversity conservation in forestry practices, and is vital for prioritizing biological conservation and management of the world's forest ecosystems (Liang et al. 2016a). However the role of biodiversity on forest productivity has not been extensively studied and remains a debated topic (Grime 1973, Al-Mufti et al. 1977, Adler et al. 2011). There exist in the literature a consensus of a *unimodal productivity-biodiversity relationship*, a hump-shaped curve with a single mode (see Rosenzweig and Abramsky 1993 and references therein), supported by the ecological theory (Abrams 1995) and empirical evidence (Waide et al. 1999, Mittelbach et al. 2001, Gillman and Wright 2006, Adler et al. 2011). However, this relationship also draws criticism because of insufficient ecosystem-wide evidence, see e.g. (Gilbert and Lechowicz 2004). In addition, most previous studies have primarily focused on non-woody species, see e.g. (Fraser et al. 2015, Grace et al. 2016), due in part to the structural complexity and long life cycle

of forests, as well as insufficient knowledge of the effects of site productivity on forest biodiversity.

Liang et al. (2016b) test the unimodal productivity-biodiversity relationship on 16 types of forest ecosystems in the United States and 3 forest types in China, using ground-sourced forest inventory data obtained from a half million ground-measured, permanent sample plots. All 19 of the forest types are found to have a consistent unimodal productivity-biodiversity relationship. Furthermore, it is found that most of these forest types exhibit a maximum tree species richness where productivity is at a medium level (Liang et al. 2016b). Liang et al. also establishes a theoretical basis for the development of a Stand Biodiversity Capacity (*SBC*) framework. It has been a common hypothesis in forestry literature that certain optimal stand structure for best growth or yield exists and varies by forest type (Ginrich 1967, Pukkala 2002, Bettinger et al. 2010). For over a century, stocking charts (e.g. Ginrich 1967) have been developed for various forest types to assist foresters and forestry field crews in achieving this optimum (Husch et al. 2003). How much biodiversity should be present within a forested ecosystem is an important question, but one that is difficult to answer due to complex ecological interactions and components of the ecosystem. *SBC* stocking charts would give forest managers a useful tool to evaluate the biodiversity of the forests they manage and create land management plans that integrate conservation into forest management strategies.

Here, to facilitate the integration of biodiversity conservation and ecological restoration into conventional forestry practices, we developed *SBC* charts for these 19 forest types in the U.S. and China to determine the optimal level of tree species diversity in each local forest stand. In addition, we created digital maps of *SBC* and current biodiversity stocking to illustrate geographic patterns of current and optimal tree species diversity across the contiguous United

States of America, Alaska, and Wangqing Forestry Bureau of China. These maps illustrate areas of high conservation potential across forest types and ecosystems and may form the basis for research projects at smaller spatial scales.

## MATERIAL AND METHODS

The ground-measured forest inventory data from the same three databases in Liang et al. (2016b) were used in this study, which contain the Forest Inventory and Analysis (FIA) of the United States (O'Connell 2014), the Cooperative Alaska Forest Inventory (CAFI, Malone et al. 2009), and the Forest Management Planning Inventory (FMPI) of Northeastern China (He et al. 2013). Data from these three sources were collected using similar sampling designs and measurement protocols (Curtis 1983). Thus, data from the three databases are comparable.

FIA data is collected throughout the forested areas of the United States as a survey of the extent and status of the nation's forests. FIA plots are permanent sample plots from which data were collected periodically. The year for which FIA data used in this study was collected varies between 1968 and 2013. Due to the passage of the Food Security Act of 1985, the geographic coordinates of FIA ground plots are changed slightly to protect landowner privacy. However, true plot locations are within 0.80 to 1.61 km of the changed values so impact is negligible (O'Connell 2014). The FIA plots are 0.04 ha in size and are placed on a hexagonal grid so that there is a one plot for every 2,428 ha of forested land (O'Connell 2014). FIA plots are distributed across the 48 contiguous U.S. states and southeastern Alaska (O'Connell 2014).

CAFI data has been collected from permanent sample plots located throughout southcentral Alaska since 1994 (Malone et al. 2009). The CAFI plots are 0.04 ha in size, square in shape, and follow similar protocol (Curtis 1983) as FIA plots. The purpose of CAFI plots is to

provide census level data of all plant species throughout southcentral Alaska on a periodic basis (Malone et al. 2009).

The FMPI collects data from permanent sample plots to assist with forest management planning design. The FMPI is conducted on government owned tree farms, parks and reserves, and county level units every 10 years (Lei 2009). FMPI data used in this study was collected in 2007 from the Wangqing Forestry Bureau, Jilin Province at the forest management unit level (He et al. 2013). FMPI plots are established across the sample area on a 1km by 2km grid. The plots are 600 m<sup>2</sup> in size and rectangular in shape.

Once data from the three sources were compiled together, attributes useful for developing the SBC charts and maps were selected using a hierarchical partitioning (HP) approach (Mac Nally and Walsh 2004) to avoid potential bias caused by multicollinearity. The attributes chosen were site productivity, species richness, stand basal area, and forest type (Table 1, Figs.1 and 2, see also Liang et al. 2016b).

<Table 1>

<Fig. 1>

<Fig. 2>

A spatially explicit index *SBC* to represent the amount of tree species that a forest stand is capable of sustaining takes the following form (Liang et al. 2016b):

$$SBC(\mathbf{s}) = E(N_i(\mathbf{s})) = \beta_{i0} + \beta_{i1} \cdot C(\mathbf{s}) + \beta_{i2} \cdot C^2(\mathbf{s}) + \beta_{i3} \cdot B(\mathbf{s}) \cdot C(\mathbf{s}) + \beta_{i4} \cdot B(\mathbf{s}), \quad (1)$$

where *SBC* can be calculated as the expected species richness of forest type *i* at point locations *s*, *N<sub>i</sub>(s)*, based on the productivity, *C(s)*, and stand basal area, *B(s)*, at that point.  $\beta_i$ 's are coefficients estimated by the geostatistical models (Liang et al. 2016b).



For a given forest type  $i$  and spatial location  $s$ ,  $SBC$  would only change with either site productivity or stand basal area. Based on this relationship, we developed  $SBC$  charts to show how  $SBC$  would respond to the changes in basal area by site productivity class. For a given site productivity  $C$  and forest type  $i$ , the intercept and slope of the linear  $SBC$  trends follow these equations:

$$\text{Intercept} = \beta_{i0} + \beta_{i1} \cdot C + \beta_{i2} \cdot C^2, \quad (2)$$

$$\text{Slope} = \beta_{i3} \cdot C + \beta_{i4}, \quad (3)$$

By aggregating the current point values of biodiversity (tree species richness) and expected  $SBC$ , we further created maps showing existing biodiversity and  $SBC$  across the contiguous United States, Alaska, and the Wangqing Forestry Bureau of China (Fig. 3). Each map consisted of 5km by 5km pixels, and each pixel represented the mean observed species richness (at 0.04-ha basis for the United States and 0.06-ha basis for China), and the mean estimated  $SBC$  of all the permanent sample plots located within this pixel.

## RESULTS

The  $SBC$  charts generally showed a positive relationship between basal area and the optimal tree species diversity with a few exceptions for the Northern Pine forest type in the United States, the Alaska boreal forest in the United States, and the Northeastern Mixed Forest type in China (Figs. 2, 3, 4, and the Supplemental Materials). The  $SBC$  stocking charts illustrated the optimal tree species richness under different levels of stand basal area and site productivity. For each site productivity level, there was a straight line outlining the expected basal area-biodiversity relationship. The intercepts and slopes varied considerably by forest type. The Pinyon/Juniper forest type, which has primarily low site productivity levels (Fig. 1), was found to have

intercepts and slopes that increased with site class (Fig. 2). The intercepts for the Pinyon/Juniper ranged from 1.913 (site class VII) to 4.934 (site class I) while the slopes of *SBC* ranged from 0.017 (site class VII) to 0.244 (site class I). This is indicative that the better the site productivity was, the higher the *SBC* will be within the Pinyon/Juniper forest type. *SBC* was highest for medium levels of site class within the broadleaf forest of northeast China (Fig. 4A). Within the broadleaf forest type of northeast China, intercepts ranged from -1.242 (site class I) to 7.661 (site class IV) and slopes ranged from -0.005 (site class VII) to 0.047 (site class I). Both the Douglas-fir and Oak/Hickory forest types generally have high levels of site productivity (Fig. 1). For both of these forest types, *SBC* for all site classes, shows a positive relationship between basal area and species richness (Fig. 4B, 4C). Higher site classes exhibited higher *SBC* than lower site classes for both the Douglas-fir and Oak/Hickory forest types. Intercepts ranged from 3.061 (site class I) to 3.873 (site class IV) and slopes ranged from 0.010 (site class I) to 0.028 (site class 7) for the Douglas-fir forest type. For the Oak/Hickory forest type, intercepts ranged from 3.406 (site class VII) to 6.408 (site class III) and slopes ranged from 0.0473 (site class I) to 0.173 (site class VII).

<Fig. 3>

<Fig. 4>

In general, the Eastern United States showed higher *SBC* than the western part of the country (Fig. 3). The Appalachian region had the highest *SBC* presumably due to a positive effect of geographic heterogeneity on biodiversity (Ricklefs and Schluter 1993). Across the United States, approximately 50 percent of forest stands were overstocked in terms of tree diversity, and 42 percent understocked. The substantially understocked forests were largely found in the southern pine region (state of Florida, Southern Georgia, and the Gulf Coast),

Central Wisconsin, Northern Minnesota, Black Hills National Forests in South Dakota, and the Pacific Northwest (states of Oregon and Washington). Although several large areas of the lower interior west (mostly located in Colorado and Utah) had moderate *SBC*, they are substantially understocked (Fig. 3). A possibility is that the distance between these forests of the lower interior west and a diverse seed source prohibits these stands from attaining full biodiversity stocking. Much of southcentral and southwestern Alaska exhibited low *SBC* and understocked to substantially understocked biodiversity. Interestingly, the Wangqing Forestry Bureau of China was found to have low *SBC* with fully stocked to overstocked biodiversity levels (Fig. 3).

## DISCUSSION

In this paper we focused on applications of the productivity-biodiversity relationship in forest management and developed the *SBC* framework to facilitate the integration of biological conservation in forestry practices. The *SBC* stocking chart provides a stocking standard in terms of species richness for conventional forestry practices. The maps we created of *SBC* and current biodiversity stocking provide a visual and spatial interpretation of *SBC* and existing biodiversity stocking levels across the United States to assist land managers, policy makers, and others with addressing conservation issues.

For each site productivity level, there was a straight line outlining the expected basal area-biodiversity relationship. Comparison of the actual tree species richness to this line would provide useful indication of *biodiversity stocking*, *i.e.* the stocking level in terms of biodiversity. We developed a four-level color-coded index as a straightforward measure of biodiversity stocking:

- Overstocked (Green):  $N \in [SBC, \infty)$ ,

- Fully stocked (Yellow):  $N \in [SBC - s.e.(SBC), SBC),$
- Understocked (Orange):  $N \in [SBC - 1.96 \cdot s.e.(SBC), SBC - s.e.(SBC)),$
- Substantially understocked (red):  $N \in [0, SBC - 1.96 \cdot s.e.(SBC)),$

where  $N$  and  $s.e.(SBC)$  were actual species richness and the standard error of  $SBC$ , respectively.

Biodiversity stocking was a stand-level measure derived from  $SBC$  and actual biodiversity.

Overstocked and fully stocked imply that actual biodiversity exceeded or approached the expected value and efforts to further improve the species richness might not be effective. A plot in the Chinese broadleaf forest type with a species richness of 5, basal area of 14.95, and a site class of 2 is representative of an overstocked stand (Fig. 4 A). A possible explanation for an overstocked stand would be that there are nonnative invasive species present. These species interfere with the natural ecology of a forest (Gordon 1998, Stinson et al. 2006, Vilà et al. 2011) and would likely contribute to a stand being overstocked in terms of species richness. Another possibility is that the evenness of native species may be changing in an overstocked stand. The abundance of some native species may be increasing in a specific forest, causing them to appear more frequently on inventory plots. This would create a situation of overstocked species richness. Whether or not measures should be taken to reduce species richness in an overstocked forest would depend on the cause of the overstocked species richness and effects that the overstocked status of species richness is having on the forest. For example, if a forest is overstocked in species richness due to an influx of invasive, nonnative species, it would be advisable to reduce the species richness to prevent native species from being outcompeted and hindered in their contribution to the forest's overall productivity.

According to our  $SBC$  chart for the Douglas-fir forest type (Fig. 4 B), a plot with a species richness of 4, basal area of 37.55, and a site class of 2, is just over the classification of

fully stocked. While management actions depend on the specific ecological properties of a given forest, in a situation such as this, it is quite likely that efforts to improve species richness would not be a viable and efficient investment, given the limited capacity of local environment in supporting a diversity tree community.

Understocked and substantially understocked suggest that such stands have good potential for sustaining a higher biodiversity. An FIA plot from our database, located within the oak/hickory forest type, having a species richness of 5, site class of 6, and a basal area of 27.52 is an example of a plot that is understocked according to our SBC chart for the oak/hickory forest type (Fig. 4 C). For natural forests in this category, biological conservation could be effective to increase the species richness, and for reclaimed forests from mining, agriculture, *etc.*, ecological restoration would be more capable of rebuilding a diverse community.

It should be mentioned however, that while the SBC charts provide a threshold for desired species richness for specific forest types, care must be taken to integrate local knowledge specific to local scale into management strategies derived from these charts. Furthermore, *SBC* is a tool set based on species richness and does not address *the choice of species* to be conserved or introduced for understocked forests. To this end, we have several existing tools to assist the selection of species in biological conservation and natural resource management, including the species' invasive and conservation status (e.g. the International Union for Conservation of Nature—IUCN Red List of Threatened Species), and Productivity Impact Index (PII) that quantifies individual species' inherent value in maintaining current ecosystem productivity (Liang et al. 2015).

## CONCLUSIONS

We developed *SBC* to represent the level of biodiversity that a forest stand is capable of sustaining. We made large-scale digital maps to illustrate the productivity and diversity of forests across the study region, and their potential of sustaining a greater biodiversity. We further developed *SBC* charts, one for each sampled forest type, to provide a more direct answer to whether or not a specific forest at a local level can sustain a higher level of biodiversity. Combined with knowledge specific at the local scale, these forest type specific *SBC* charts will facilitate integration of biodiversity conservation and ecological restoration into forestry practices.

## COMPETING INTERESTS

The authors declare that they have no competing interests.

## AUTHORS' CONTRIBUTION

MZ assisted with designing the concept of *SBC* and helped to draft the manuscript. JL compiled data, performed data analysis, assisted with designing the concept of *SBC*, and helped to draft the manuscript. JW performed data analysis, compiled data, and assisted with drafting the manuscript. XL assisted with the compilation of data, performed data analysis, and helped to draft the manuscript. All authors read and approved the final manuscript.

## ACKNOWLEDGEMENTS

This study is supported in parts by the West Virginia University, and the USDA McIntire-Stennis Funds WVA00126. We thank the Global Forest Biodiversity Initiative (GFBI) for establishing the data standards and collaborative framework.

## LITERATURE CITED

- Abrams, P. A. 1995. Monotonic or unimodal diversity-productivity gradients: what does competition theory predict? *Ecology* **76**:2019-2027.
- Adler, P. B., E. W. Seabloom, E. T. Borer, H. Hillebrand, Y. Hautier, A. Hector, W. S. Harpole, L. R. O'Halloran, J. B. Grace, T. M. Anderson, J. D. Bakker, L. A. Biederman, C. S. Brown, Y. M. Buckley, L. B. Calabrese, C.-J. Chu, E. E. Cleland, S. L. Collins, K. L. Cottingham, M. J. Crawley, E. I. Damschen, K. F. Davies, N. M. DeCrappeo, P. A. Fay, J. Firn, P. Frater, E. I. Gasarch, D. S. Gruner, N. Hagenah, J. Hille Ris Lambers, H. Humphries, V. L. Jin, A. D. Kay, K. P. Kirkman, J. A. Klein, J. M. H. Knops, K. J. La Pierre, J. G. Lambrinos, W. Li, A. S. MacDougall, R. L. McCulley, B. A. Melbourne, C. E. Mitchell, J. L. Moore, J. W. Morgan, B. Mortensen, J. L. Orrock, S. M. Prober, D. A. Pyke, A. C. Risch, M. Schuetz, M. D. Smith, C. J. Stevens, L. L. Sullivan, G. Wang, P. D. Wragg, J. P. Wright, and L. H. Yang. 2011. Productivity Is a Poor Predictor of Plant Species Richness. *Science* **333**:1750-1753.
- Al-Mufti, M., C. Sydes, S. Furness, J. Grime, and S. Band. 1977. A quantitative analysis of shoot phenology and dominance in herbaceous vegetation. *The Journal of Ecology*:759-791.
- Bettinger, P., K. Boston, J. P. Siry, and D. L. Grebner. 2010. *Forest management and planning*. Academic Press, Burlington, MA.
- Butchart, S. H. M., M. Walpole, B. Collen, A. van Strien, J. P. W. Scharlemann, R. E. A. Almond, J. E. M. Baillie, B. Bomhard, C. Brown, J. Bruno, K. E. Carpenter, G. M. Carr, J. Chanson, A. M. Chenery, J. Csirke, N. C. Davidson, F. Dentener, M. Foster, A. Galli, J. N. Galloway, P. Genovesi, R. D. Gregory, M. Hockings, V. Kapos, J.-F. Lamarque, F. Leverington, J. Loh, M. A. McGeoch, L. McRae, A. Minasyan, M. H. Morcillo, T. E. E. Oldfield, D. Pauly, S. Quader, C. Revenga, J. R. Sauer, B. Skolnik, D. Spear, D. Stanwell-Smith, S. N. Stuart, A. Symes, M. Tierney, T. D. Tyrrell, J.-C. Vié, and R. Watson. 2010. Global Biodiversity: Indicators of Recent Declines. *Science* **328**:1164-1168.
- Curtis, R. O. 1983. Procedures for establishing and maintaining permanent plots for silvicultural and yield research. Gen. Tech. Rep. PNW-155, U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station, Portland, OR.
- FAO. 2011. *Assessing forest degradation: towards the development of globally applicable guidelines* United Nations Food and Agriculture Organization, Rome, Italy.

- FAO. 2015. Global Forest Resources Assessment 2015 - How are the world's forests changing?. Food and Agriculture Organization of the United Nations.
- Fraser, L. H., J. Pither, A. Jentsch, M. Sternberg, M. Zobel, D. Askarizadeh, S. Bartha, C. Beierkuhnlein, J. A. Bennett, and A. Bittel. 2015. Worldwide evidence of a unimodal relationship between productivity and plant species richness. *science* **349**:302-305.
- Gilbert, B. and M. J. Lechowicz. 2004. Neutrality, niches, and dispersal in a temperate forest understory. *Proceedings of the National Academy of Sciences* **101**:7651-7656.
- Gillman, L. N. and S. D. Wright. 2006. The influence of productivity on the species richness of plants: a critical assessment. *Ecology* **87**:1234-1243.
- Ginrich, S. F. 1967. Measuring and evaluating stocking and stand density in upland hardwood forests in the Central States. *Forest Science* **13**:38-53.
- Gordon, D. R. 1998. Effects of invasive, non-indigenous plant species on ecosystem processes: lessons from Florida. *Ecological Applications* **8**:975-989.
- Grace, J. B., T. M. Anderson, E. W. Seabloom, E. T. Borer, P. B. Adler, W. S. Harpole, Y. Hautier, H. Hillebrand, E. M. Lind, and M. Pärtel. 2016. Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* **529**:390-393.
- Grime, J. P. 1973. Competitive exclusion in herbaceous vegetation. *Nature* **242**:344-347.
- He, P., H. Zhang, X. Lei, and X. Li. 2013. Estimation of spatial distribution of tree species diversity based on Universal Kriging Model. *Journal of Central-South University of Forestry & Technology* **33**:67-71.
- Husch, B., T. W. Beers, and J. A. Kershaw Jr. 2003. Forest mensuration. 4th edition. John Wiley & Sons, Hoboken, New Jersey
- Lei, X., M. Tang, Y. Lu, L. Hong, and D. Tian. 2009. Forest inventory in China: status and challenges. *International Forestry Review* **11**:52-63.
- Liang J, Crowther TW, Picard N, Wiser S, Zhou M, Alberti G, Schulze E-D, McGuire AD, Bozzato F, Pretzsch H, de-Miguel S, Paquette A, Hérault B, Scherer-Lorenzen M, Barrett CB, Glick HB, Hengeveld GM, Nabuurs G-J, Pfautsch S, Viana H, Vibrans AC, Ammer C, Schall P, Verbyla D, Tchebakova N, Fischer M, Watson JV, Chen HYH, Lei X, Schelhaas M-J, Lu H, Gianelle D, Parfenova EI, Salas C, Lee E, Lee B, Kim HS, Bruelheide H, Coomes DA, Piotto D, Sunderland T, Schmid B, Gourlet-Fleury S, Sonké B, Tavani R, Zhu J, Brandl S, Vayreda J, Kitahara F, Searle EB, Neldner VJ, Ngugi MR, Baraloto C, Frizzera L, Bałazy R, Oleksyn J, Zawila-Niedzwiecki T, Bouriaud O, Bussotti F, Finér L, Jaroszewicz B, Jucker T, Valladares F, Jagodzinski AM, Peri PL, Gonmadje C, Marthy W, O'Brien T, Martin EH, Marshall AR, Rovero F, Bitariho R, Niklaus PA, Alvarez-Loayza P, Chamuya N, Valencia R, Mortier F, Wortel V, Engone-Obiang NL, Ferreira LV, Odeke DE, Vasquez RM, Lewis SL, Reich PB. 2016a. Positive biodiversity-productivity relationship predominant in global forests. *Science* **354** (6309). doi:10.1126/science.aaf8957
- Liang, J., J. V. Watson, M. Zhou, and X. Lei. 2016b. Effects of productivity on biodiversity in forest ecosystems across the United States and China. *Conservation Biology* **30**:308-317.
- Liang, J., M. Zhou, P. C. Tobin, A. D. McGuire, and P. B. Reich. 2015. Biodiversity influences plant productivity through niche-efficiency. *Proceedings of the National Academy of Sciences* **112**:5738-5743.
- Lindenmayer, D. B., J. F. Franklin, and J. Fischer. 2006. General management principles and a checklist of strategies to guide forest biodiversity conservation. *Biological Conservation* **131**:433-445.



- Mac Nally, R. and C. J. Walsh. 2004. Hierarchical partitioning public-domain software. *Biodiversity & Conservation* **13**:659-660.
- Malone, T., J. Liang, and E. C. Packee. 2009. Cooperative Alaska Forest Inventory. Gen. Tech. Rep. PNW-GTR-785, USDA Forest Service, Pacific Northwest Research Station, Portland, OR.
- Mittelbach, G. G., C. F. Steiner, S. M. Scheiner, K. L. Gross, H. L. Reynolds, R. B. Waide, M. R. Willig, S. I. Dodson, and L. Gough. 2001. What is the observed relationship between species richness and productivity? *Ecology* **82**:2381-2396.
- O'Connell, B. M., E. B. LaPoint, J. A. Turner, T. Ridley, S. A. Pugh, A. M. Wilson, K. L. Waddell, and B. L. Conkling. 2014. The Forest Inventory and Analysis Database: Database description and user guide version 6.0.1 for Phase 2. U.S. Department of Agriculture, Forest Service., Washington, D.C.
- Pan, Y., R. A. Birdsey, O. L. Phillips, and R. B. Jackson. 2013. The structure, distribution, and biomass of the world's forests. *Annual Review of Ecology, Evolution, and Systematics* **44**:593-622.
- Pukkala, T. 2002. Multi-objective forest planning. Kluwer Academic Publishers, Dordrecht, Netherlands.
- Ricklefs, R. E. and D. Schluter. 1993. Species diversity in ecological communities: historical and geographical perspectives. University of Chicago Press, Chicago, IL.
- Rosenzweig, M. L. and Z. Abramsky. 1993. How are diversity and productivity related? Pages 52-65 *in* R. E. Ricklefs and D. Schluter, editors. *Species Diversity in Ecological Communities*. The University of Chicago Press, Chicago, IL.
- Stinson, K. A., S. A. Campbell, J. R. Powell, B. E. Wolfe, R. M. Callaway, G. C. Thelen, S. G. Hallett, D. Prati, and J. N. Klironomos. 2006. Invasive plant suppresses the growth of native tree seedlings by disrupting belowground mutualisms. *PLoS biology* **4**:e140.
- Vilà, M., J. L. Espinar, M. Hejda, P. E. Hulme, V. Jarošík, J. L. Maron, J. Pergl, U. Schaffner, Y. Sun, and P. Pyšek. 2011. Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecology letters* **14**:702-708.
- Waide, R., M. Willig, C. Steiner, G. Mittelbach, L. Gough, S. Dodson, G. Juday, and R. Parmenter. 1999. The relationship between productivity and species richness. *Annual Review of Ecology and Systematics*:257-300.

## Tables

Table 1. Key attribute definitions and units

	Units	Name	Definition		
<i>C</i>	$\text{m}^3\text{ha}^{-1}\text{yr}^{-1}$	Site productivity	Potential increase of timber volume in mean annual increment (MAI) calculated from site class (below) and physiographic conditions		
			Site class	Corresponding Productivity Range	Mean Productivity
				$(\text{m}^3\text{ha}^{-1}\text{yr}^{-1})$	$(\text{m}^3\text{ha}^{-1}\text{yr}^{-1})$
			I	[15.74, $+\infty$ )	16.00
			II	[11.55, 15.74)	13.65
			III	[8.40, 11.55)	9.98
			IV	[5.95, 8.40)	7.18
			V	[3.50, 5.95)	4.73
			VI	[1.40, 3.50)	2.45
VII	[0, 1.40)	0.70			
<i>N</i>		Species richness	The total number of different tree species present on the sampling plot		
<i>B</i>	$\text{m}^2\text{ha}^{-1}$	Stand basal area	Total cross-sectional area at breast height (DBH) of all living trees, used as a substitute for resource acquisition and intra- and inter-specific competition		
<i>i</i>		Forest type	A forest type represents a distinguishable forest ecosystem having relatively homogeneous species composition, geographic features, and environmental conditions		

## FIGURE TITLES

Figure 1. Site productivity ( $\text{m}^3\text{ha}^{-1}\text{year}^{-1}$ , A), species richness (B), and forest type (C) across the contiguous United States and Alaska. The data were collected from 476,677 Forest Inventory and Analysis (FIA) and 785 Cooperative Alaska Forest Inventory (CAFI) sample plots (a section of the database is shown above). This figure was reproduced from Liang et al. 2016b with permission from the publisher.

Figure 2. Forest types of 1,387 forest sample plots from northeastern China. This figure was reproduced from Liang et al. 2016b with permission from the publisher.

Figure 3. Stand biodiversity capacity (*SBC*, upper figure) and biodiversity stocking (lower figure) of forest types across the contiguous United States and Alaska, and northeastern China (Insets), based on the ground-measured forest inventory data. On top is the *SBC* Charts for Pinyon/Juniper forest type in the United States. Each line shows the relationship between basal area and expected *SBC* for one site class (Class I is the most productive and Class VII the least productive, see Table 1 for details). *SBC* charts for additional forest types are shown in the supplemental material.

Figure 4. *SBC* Charts for the Northeastern Broadleaf forest type in China, the Douglas-fir forest type of the western United States, and the Oak/Hickory forest type of the eastern United States. Each line shows the relationship between basal area and expected *SBC* for one site class (Class I is the most productive and Class VII the least productive, see Table 1 for details). *SBC* charts for additional forest types are provided in the supplemental materials.

## Figures

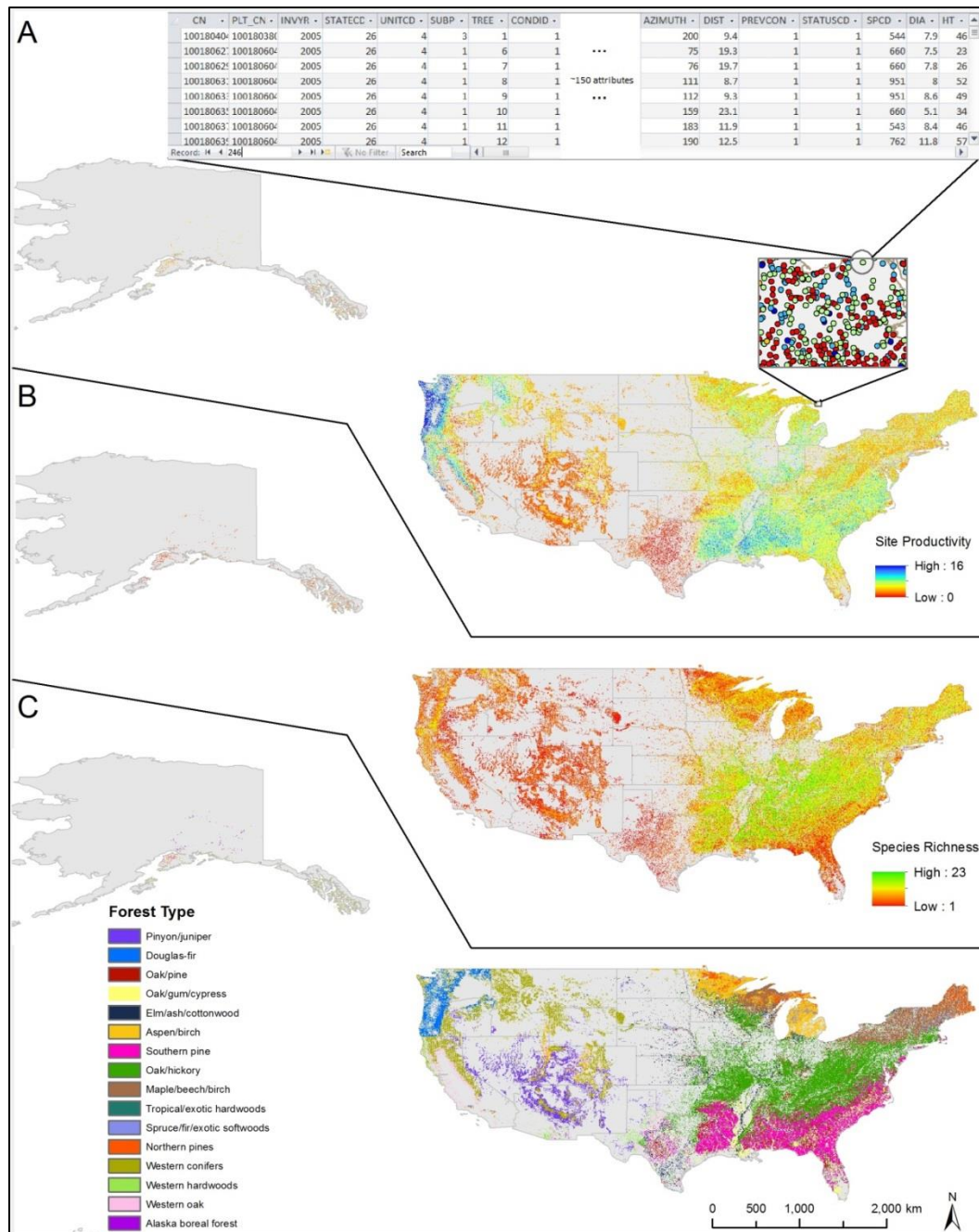


Figure 1. Site productivity ( $\text{m}^3\text{ha}^{-1}\text{year}^{-1}$ , A), species richness (B), and forest type (C) across the contiguous United States and Alaska. The data were collected from 476,677 Forest Inventory and Analysis (FIA) and 785 Cooperative Alaska Forest Inventory (CAFI) sample plots (a section of the database is shown above). This figure was reproduced from Liang et al. 2016b with permission from the publisher.

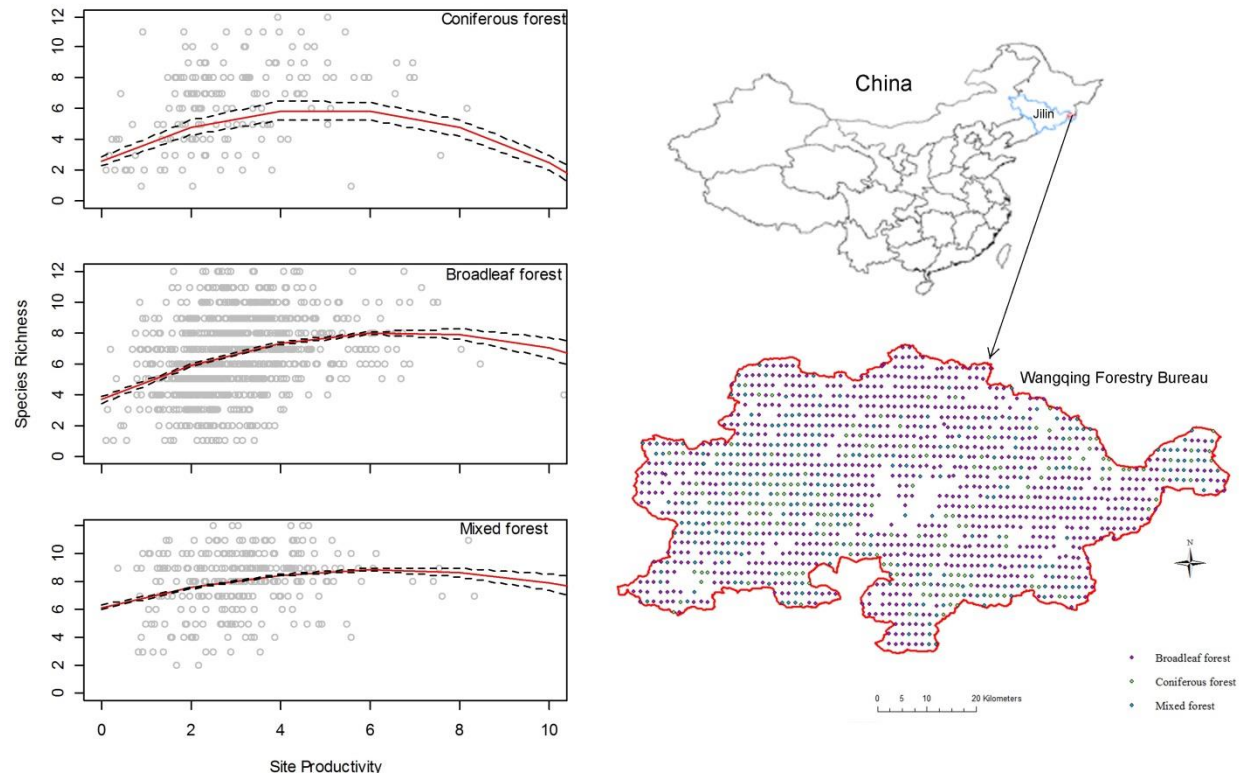


Figure 2. Forest types of 1,387 forest sample plots from northeastern China. This figure was reproduced from Liang et al. 2016b with permission from the publisher.

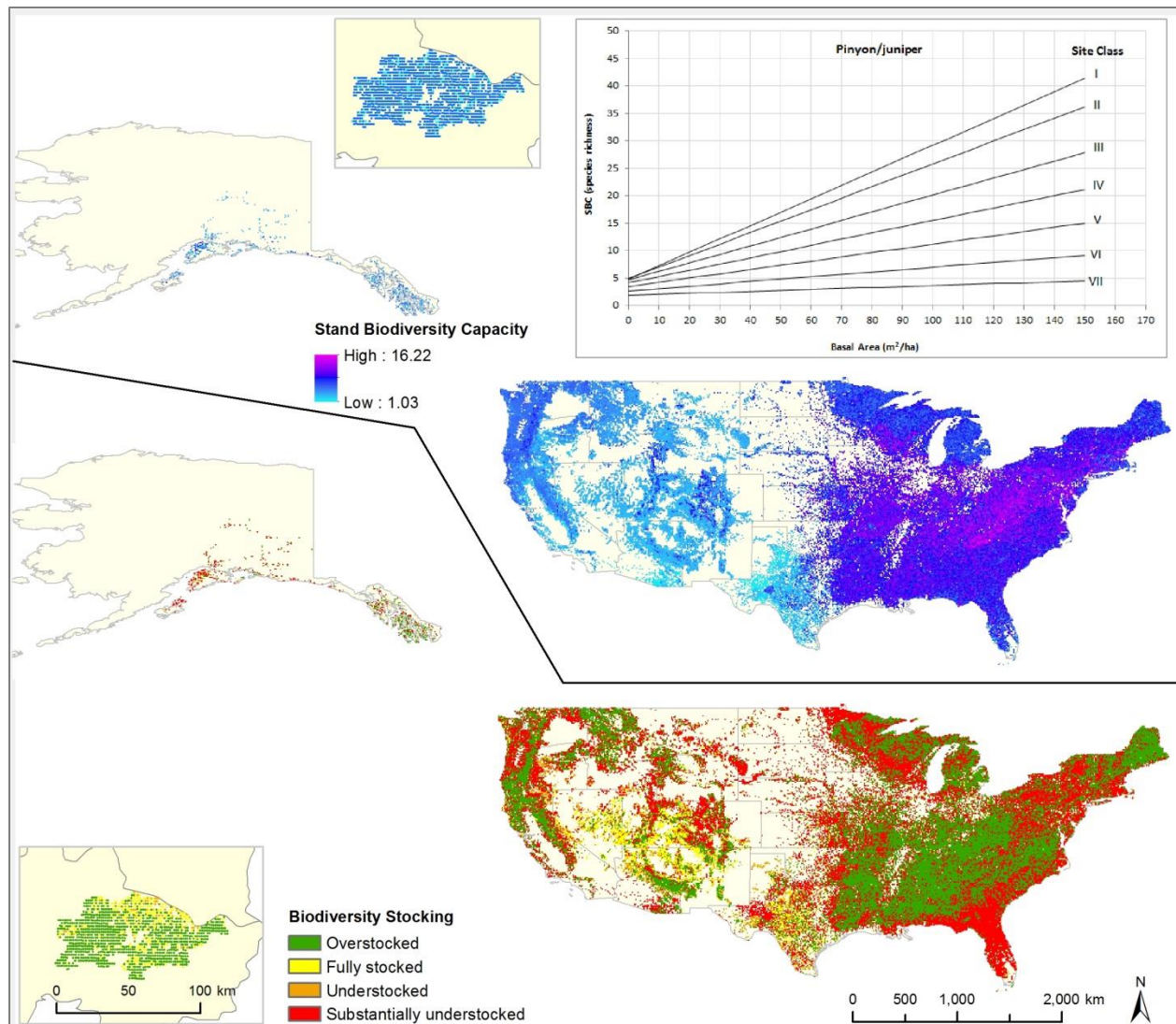


Figure 3. Stand biodiversity capacity (*SBC*, upper figure) and biodiversity stocking (lower figure) of forest types across the contiguous United States and Alaska, and northeastern China (Insets), based on the ground-measured forest inventory data. On top is the *SBC* Charts for Pinyon/Juniper forest type in the United States. Each line shows the relationship between basal area and expected *SBC* for one site class (Class I is the most productive and Class VII the least productive, see Table 1 for details). *SBC* charts for additional forest types are shown in the supplemental material.

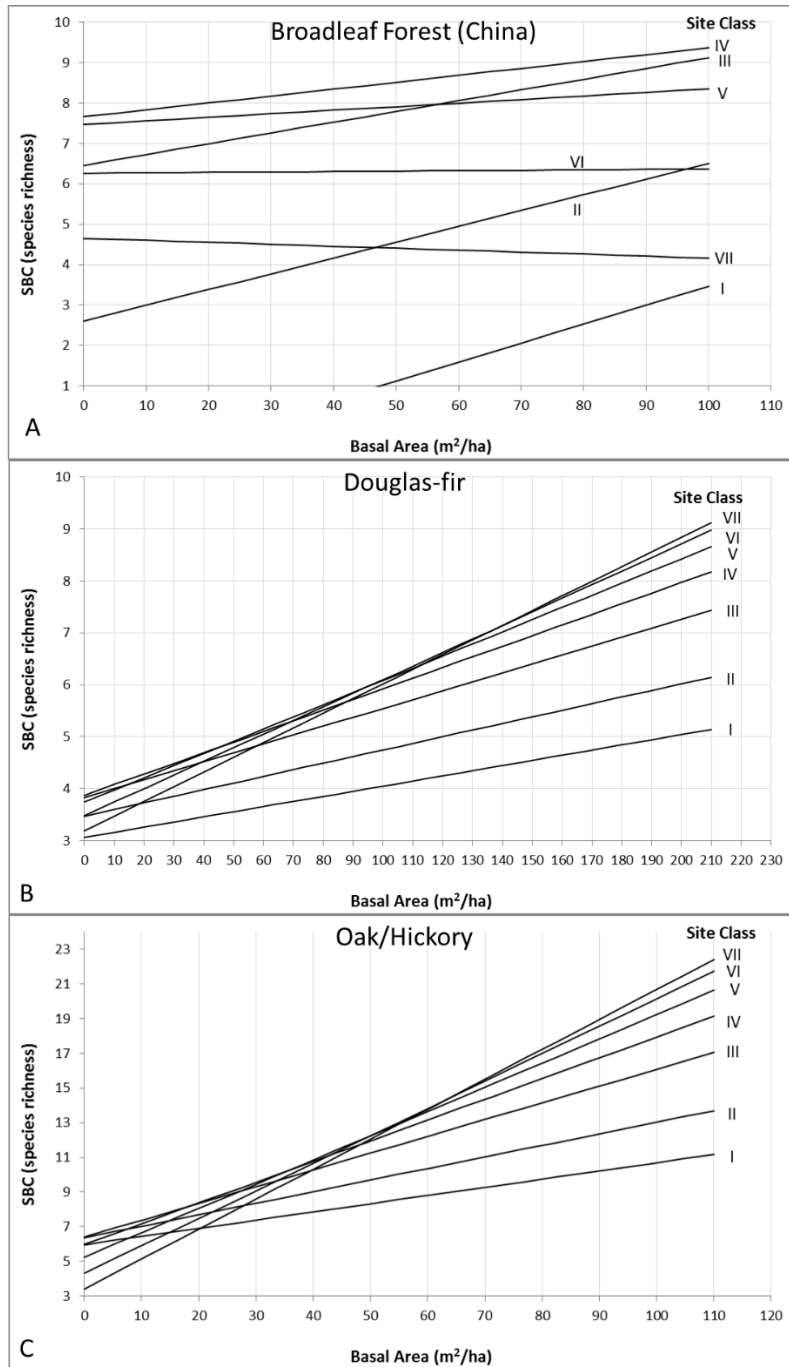


Figure 4. *SBC* Charts for the Northeastern Broadleaf forest type in China, the Douglas-fir forest type of the western United States, and the Oak/Hickory forest type of the eastern United States. Each line shows the relationship between basal area and expected *SBC* for one site class (Class I is the most productive and Class VII the least productive, see Table 1 for details). *SBC* charts for additional forest types are provided in the supplemental materials.

## Chapter 3:

### Evaluation of Empirical Dynamic Modeling in a Hydrologic System in West Virginia

#### **Abstract**

Environmental studies often deal with dynamical systems exhibiting changes in environmental factors and their relationships over time. Most of these studies assume that such systems are linear and employ classic time-series approaches for modeling factors and detecting the direction and strength of the relationships. Empirical dynamic modeling (EDM) is a recently developed method specifically for modeling nonlinear dynamical systems, effective for situations when linear approaches fail, such as mirage correlation. Using 29- year stream discharge, precipitation, air temperature, vapor pressure deficit, and normalized difference vegetation index (NDVI) data from the state of West Virginia, we contrast a few classic linear methods which include uni- and multi-variate modeling and Granger Causality test, to EDM methods which include simplex models, s-mapping, and Convergent Cross Mapping (CCM). Our results show that EDM provides more predictive power over linear methods with all data lengths, temporal frequencies, and number of variables. Additionally, EDM allows for the analysis of interactive strength of variables across the state space, which could help inform and refine process-based hydrologic models. These results highlight the theoretical strengths and benefits of using EDM, over linear approaches. The empirical results suggest that the process of convergent cross-mapping should be used to supplement the Granger causality test when determining causality among variables of watersheds in northern West Virginia and areas of similar climatic conditions, ecosystem functioning processes, and geography.



## Introduction

Dynamic behavior refers to the situation where behavior of a variable or a group of interacting variables is dependent upon the state of another variable or the overall state of the system, i.e., state-dependent behavior. Classically, linear time series methods have been applied to study complex and dynamical systems. For example, autoregressive integrated moving average (ARIMA) models have largely been used for univariate and multivariate time series data. Granger Causality testing is a standard procedure for the significance, magnitude, and direction of bilateral relationships. These methods, all based on linear relationships between variables, are often ineffective or even fail when systems exhibit complex nonlinear dynamic behavior (Chang et al. 2017). For example, it is difficult to detect the relationship between two variables with the Granger causality test if their relationship varies with the state of the system, i.e., they display mirage correlation (Sugihara et al. 2012).

To address the need of properly analyzing data from a system comprised of state-dependent nonlinear behavior, in recent decades, statistical methods have been developed that assume nonlinearity of time series data (Anderson et al. 2008; Chang et al. 2017; Glaser et al. 2014; Sugihara and May 1990; Ye et al. 2015). EDM is an empirically driven nonparametric method designed to leverage the nonlinear dynamic information that linear models smooth over (Sugihara et al. 2012; Ye et al. 2015). Specifically, EDM can describe the changes of a system's general trajectory in the state space at any given time step, a direct result of the systems' dynamic behavior. Additionally, EDM can describe when the direction (+/-) and strength of the correlations vary with time, which is a result of the change of interactions between variables as the state of the system alters (Chang et al. 2017; Sugihara et al. 2012; Ye et al. 2015). Unlike many traditional statistical approaches that attempt to fit equations under certain hypotheses,

EDM relies only upon the empirical data to determine these relationships, thus the complex interaction. Consequently, EDM is not subject to the constraints of hypotheses but rather to the abundance and nature of time series data (Ye et al. 2015).

EDM is centered around the process of mapping the state space of a dynamic system, which is dependent on the system's complexity, within multi-dimensional space. The complexity of a dynamic system may be defined as the number of signals, or Embedding dimensions (E), within a single time series data (Ye 2017). When implementing EDM, the state space of a system is generated by continually mapping the multiple time series of the system onto the multi-dimensional space. This results in an E-dimensional map of the contained behavior of the system under all of its observed states; hence this generates a model of the systems' state space in multi-dimensional space. Conversely, we can describe the process in reverse, demonstrating the ability of EDM to forecast. Starting with the unobservable state space of a system we can think of EDM predictions as projections of the state space model on to a single axis of the state space model that represents the variable(s) in the model (Ye 2017).

A particular advantage of EDM is that it allows for the behavior of a dynamical system involving multiple variables to be determined from a single time series, based on Taken's theorem (1981). Simply put, this theorem proves that all the information of a system can be extracted from a single time series data of that system by using its time-delayed embeddings to reconstruct the system's state space. In other words, by reconstructing the state space of a system through lags of univariate time series, the behavior of a combination of variables in this system can be obtained (Takens 1981). This advantage has tremendous appeal to empirical studies of ecological studies because in many situations, there is only one variable observable for a prolonged period, resulting in a single time series. In addition, EDM incorporates a nonlinear

counterpart of Granger causality test, convergent cross mapping (CCM). CCM can identify more subtle relationships such as correlation without causation, and the effects of causation when it changes in direction and strength over time (Sugihara et al. 2012; Ye et al. 2015).

When studying at the ecosystem level, scientists often attempt to understand and model the processes that comprise the functioning of an ecosystem. These processes typically exhibit nonlinearity and dynamic behavior. Thus, EDM has potential to be a very useful tool for the purpose of analyzing Ecosystem functioning characteristics and processes (Ye 2017; Ye et al. 2015).

A hydrologic system is an excellent example of an ecosystem process exhibiting nonlinearity and dynamic behavior (KUNDZEWICZ and NAPIÓRKOWSKI 1986). In this study, a hydrologic system located in northern West Virginia serves as a case study for demonstrating the efficiency of EDM as a tool to model nonlinear, dynamic processes comprising ecosystem functioning. The goal of this study was to address three primary objectives.

The first objective of this study was to determine the efficiency of EDM under varying data resolution and time series length of environmental data. A secondary objective is to evaluate EDM as an alternative and supplementary method to the Granger causality test and ARIMA models. Lastly, we aimed to determine the driving variables, and their dynamic interactions, in a hydrologic system in northern West Virginia.

## Data and Methods

### Data

The study site is located in Randolph County, in northern West Virginia, at Valley Bend, on the Tygart River (Figure 1). The Tygart River drainage area is a Hydrologic Unit Code (HUC) 8. The climate is characterized by cold winters and warm, humid summers. Although summer is typically the driest season, large amounts of precipitation are possible at any time of the year (NOAA 2018).

The variables used in this study include stream discharge, daily precipitation, mean daily air temperature, vapor pressure deficit, and Normalized Difference Vegetation Index (NDVI). Stream discharge (Table 1) represents the discharge from the headwaters of the Tygart River and was downloaded from the USGS Water Resources site (<https://waterdata.usgs.gov/wv/nwis/rt>, last accessed February 28, 2018). The stream discharge data was structured in time steps of one hour from 1996 to 2017. In the year 2015, a stream discharge data set from the USGS Water Resources website that represented 12 month time sets from 1986 to 2013 for the Valley Bend site was downloaded. After reducing the hourly data to monthly data, the previous monthly dataset was combined to the newer hourly dataset to create a new dataset of stream discharge at monthly time steps from 1986 to 2017, representing a 29-year data set. The newer stream discharge data that was collected at hourly time steps was further reduced to bi-daily and daily time intervals spanning from 1996 to 2017.

The atmospheric variables of Daily precipitation, mean daily air temperature, and vapor pressure deficit were collected from the PRISM climate group website (<http://prism.oregonstate.edu/>, last accessed March 2, 2018) at daily time steps from 1996 to

2017. The variables collected from PRISM are interpolated over the rectangular area that contained the USGS stream data collection site (Daly 1994).

NDVI was collected from the USGS Landsat Missions website (<https://nex.nasa.gov/nex/projects/1349/>, last accessed March 3, 2018), for the period of 1996 to 2017, originally at two week time steps (Pinzon and Tucker 2014). The data was interpolated over the same location and also interpolated to daily time steps that were used alongside the atmospheric time series as predictor variables for Stream discharge.

### Stationary test

The critical initial step of analysis is to test for data stationarity to avoid spurious relationships, for both linear and nonlinear approaches. Stationarity refers to the assumption that the mean, variance, and any autocorrelation of each time series does not change with time. All series at their original levels were tested for stationarity using the Dickey fuller unit root test (Dickey and Fuller 1979), (Table 2). All variables except for precipitation were nonstationary (Table 2). For consistency, we took the first difference of all the variables and applied Dickey fuller unit root test again. They were found to be stationary (Table 2) and were thus used in the following analysis.

### ARIMA Models

ARIMA models forecasting stream discharge were created from monthly stream discharge data at both 19 year and 29 year time spans. Daily and bi-daily stream discharge data for a 19 year time span were also used to construct ARIMA models which predicted stream discharge.

ARIMA models employ a linear forecasting equation in which the predictors in the model are lags of both the dependent variable and errors (Nau 2018). Optimum number of lags of

autoregressive and moving average terms were determined based on Akaike's 'An Information Criterion (AIC) in order to build the ARIMA models. The ARIMA models we used took the basic form of:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

where  $p$  is the number of autoregressive terms,  $q$  is the number of lagged forecast errors in the prediction equation,  $\hat{y}$  is stream discharge at time  $t$ ,  $\phi$ 's and  $\theta$ 's are coefficients of autoregressive and moving average terms, and  $e$ 's are errors.

Fitted values for the ARIMA models represent one step predictions across the entire data set. The correlation coefficient of the second half of the fitted values of the ARIMA models was calculated using the second half of the first differenced data, as a representation of the models predictive skill under 'hold out' conditions. This validation procedure will henceforth be referred to as "ARIMA 50%".

Additionally, a cross validation method that uses the rolling window method was completed as a robust method for cross validation that is similar to the leave-one-out-method. This cross validation method used a two year training window with a one year prediction horizon for computing errors. The forecasted values for the cross validation were compared with original data that had the first difference calculated to generate a correlation coefficient. This validation procedure will henceforth be referenced as "ARIMA CV".

### Multivariate ARIMA Model

The Multivariate ARIMA model was constructed using daily observations of all four predictor variables:

$$\hat{y}_t - \phi_1 \hat{y}_{t-1} = \mu - \theta_1 e_{t-1} + \beta_1(\text{Precip}_t - \phi_1 \text{Precip}_{t-1}) + \beta_2(\text{Airtemp}_t - \phi_1 \text{Airtemp}_{t-1}) + \beta_3(\text{VPD}_t - \phi_1 \text{VPD}_{t-1}) + \beta_4(\text{NDVI}_t - \phi_1 \text{NDVI}_{t-1})$$

where  $\hat{y}$  is stream discharge at time  $t$ ,  $\theta$ 's are moving averages,  $\phi$ 's are slope coefficients,  $\mu$  is the overall average, and the  $e$  is noise error.

The optimum autoregressive and moving average terms were determined via AIC for each of the predictor variables and stream discharge to construct the multivariate ARIMA model. Once constructed, the model was validated by generating one-step forecasts for the second half of the dataset and then calculating the correlation coefficient of the predictions with the corresponding original data.

### Granger Causality

The Granger causality test was used to determine causality between stream discharge and each of the four predictor variables. A particular variable, is said to “Granger cause” another variable if the predictability of one variable decreases when the other variable is removed from the space of all possible causative variables (Sugihara et al. 2012). In the context of this analysis, causation between stream discharge and predictor variables is defined as stream discharge granger causing each of the four predictor variables as well as each of the four predictor variables granger causing stream discharge. The null hypothesis for each set of causation analysis is that there exists no granger causality between the variables. The general form of the test is:

$$y_t = \alpha + \sum_{i=1}^k \beta_i y_{t-i} + e_t \quad (\text{reduced model (R)})$$

$$y_t = \alpha + \sum_{i=1}^k \beta_i y_{t-i} + \sum_{j=1}^p \gamma_j x_{t-j} + e_t \quad (\text{full model (F)})$$

$$H_0: \gamma_1 = \gamma_2 = \dots = \gamma_k = 0 \quad (\text{i.e., the reduced model})$$

The F statistic is thus:

$$F = \frac{(SSE(R) - SSE(F))/(df_R - df_F)}{SSE(F)/df_F}$$

where,  $df_R = n - k - 1$  and  $df_F = n - k - p - 1$ . If rejected, there is no causality of variable  $x$  on variable  $y$ . The causality test of variable  $y$  on  $x$  take the symmetrical form as above.

### EDM Models

EDM is based on the assumption that the dynamical system is nonlinear. We used the s-mapping (sequential locally weighted global linear map) methods (Sugihara 1994) to test for nonlinearity. All data used met both assumptions, without further transformation. All EDMs optimized the number of embedding dimensions with predictive skill; while theta values were optimized for mean squared error when creating models for the comparison between univariate and multivariate models of daily discharge data theta values. We used two methods for EDM validation that are analogous to the validation used for the corresponding ARIMA models. First we did not specify training, and predictive libraries when making the models. This was done to force EDM to use the default leave-one-out method, and calculate the correlation coefficient across all observations and predictions. This technique is henceforth referenced as “EDM CV”. Second, we specified training libraries as the first half of the data set and prediction libraries as the second half of the data set. This method is analogous to the single-time step predictions used for ARIMA models. This technique is henceforth referenced as “EDM 50%”. Time-lagged values of stream discharge were used to construct the univariate model while multiple variables were used to construct the multivariate model. Lastly, a multiview model which used all four predictive variables along with a single lag of each variable was constructed. After testing each combination of given variables using a defined number of embedding dimensions, the multiview model then selects the best top composition of lags and variables to then combine into a single



model by averaging nearest neighbor coefficients among the best top models. Interaction strength plots were generated using the s-mapping method with no nearest neighbors, forcing the model to output s-mapping coefficients to contain the maximum amount of information in the data. Through this process, partial derivatives of all points in time for each variables are calculated and can be interpreted as interspecific interactions between all driving variables.

### Convergent Cross Mapping

The cross-mapping algorithm developed by Sugihara et al. (2012) was used to test for causation between stream discharge and each of the other four variables at daily intervals (Sugihara et al. 2012). This algorithm uses the time lags of a variable to predict the current state of another variable. If two variables are causally linked, then the resulting cross-mapping will be convergent. Here, the term convergent means that the cross-mapping skill will be improved with increasing size of data (library size) indicating causality (Chang et al. 2017). We optimized the theta value for the lowest mean square error value for each cross convergent mapping performed, while also using 1000 different libraries of different sizes for cross-mapping of one variable to the predictor variable. This allowed us to calculate the 1st and 3rd quartiles of the correlation coefficient of the observed and predicted values over a range of library sizes.

### Software

R-programming was used for all data analysis and modeling used in this study. A few packages within the R-programming software that were used are of note. ARIMA models were created using package forecast while the MARIMA model was created using the marima package. All EDM and convergent cross-mapping tasks were created using the rEDM package.

## Results

Figures 3 and 4 show the forecasting performance of stream discharge for both linear models (ARIMA) and EDM. The ability of EDM to predict stream discharge outperformed that of ARIMA in every resolution and length, regardless of the type of validation used (Fig. 1, 2). In regards to the length of time series data, the performance of ARIMA models increased with an expansion of the length of time series data. (Fig.1). However, just the opposite was observed for EDM as forecasting performance was higher for time series length of 19 years than a time series length of 29 years (Fig.1). For the processes of validation, the cross-validation method yielded models of lower forecasting ability compared to models based on the hold-out method of validation for both ARIMA models and EDM, regardless of time series length (Fig. 1). Figure 2 shows the change in model performance as resolution is increased for both ARIMA models and EDM. As data resolution was increased from monthly to bidaily, forecasting performance improved consistently for the ARIMA models regardless of validation method. Between the two methods of validation, the hold-out method yielded models of better performance than the cross-validation method for ARIMA models of all data resolutions tested (Fig. 2).

Results of EDM performance across varying data resolutions were less straightforward than those of ARIMA models. One interesting observation is that as data resolution is increased, EDM performance does not increase consistently as is the case for the ARIMA models. For both types of validation, EDM performance was lower for bimonthly time series data than for monthly time series data. Also, although the cross-validation method for daily time series data yielded a model exhibiting higher performance than the model created from the cross-validation method for monthly time series data, the hold-out method of validation for daily time series data

created a model with a lower level of performance than the model produced through the hold-out method for monthly time series data (Fig. 2). It should be noted however that for both methods of validation, bidaily time series data yielded models having the highest performance of all models analyzed. The hold-out method of validation created a model with higher forecast performance than the cross-validation method, with the exception of when daily time series data was used (Fig. 2).

Figure 3 shows a comparison of the performance between the multivariate ARIMA model and a univariate EDM, multivariate EDM, and multiview EDM. ARIMA had a lower correlation coefficient than each type of EDM. The multiview EDM, which results from the process of selecting the best models produced from EDM protocol, had the highest ability to predict stream discharge.

The results of using convergent cross mapping to identify the causality between stream discharge and the variables of precipitation, air temperature, vapor pressure, and NDVI are shown in Figure 4. It should be noted that cross-mapping is in the opposite direction of cause-effect (Chang 2017). For example, the text “Discharge xmap Precip” should be interpreted as whether or not precipitation causes stream discharge. In these figures, the grey dotted line represents the overall correlation, which is significant for all tests (Fig. 4). Furthermore the overall correlation coefficient was negative for all variables except precipitation, and for convenience is shown as being positive in these figures. The solid lines represent median of predictive skill by library size and if convergent, that is to say they increase across all library size increases, we assume that there exists a causal relationship of one variable on the other. The first and third quantiles of predictive skills of the sets of subsampled library sizes are represented by the red and blue dotted lines (Fig. 4). Figure 4 shows that between the variables of stream

discharge and precipitation, stream discharge has no significant causation on precipitation, yet precipitation has notable causation on stream discharge across all library sizes. For the other causality tests using air temperature along with VPD (i.e. stream discharge vs. air temperature, stream discharge vs. vapor pressure,) there is significantly less predictive ability when the discharge is assumed to be the driving variable. Additionally, in both of these causality tests, we see that when discharge is the causal variable there is no convergence across all library sizes tested; otherwise when air temperature, VPD, is the causal variable, convergence is seen across all library sizes tested. Finally, when we look at the causal relationships between NDVI and discharge, as well as discharge and NDVI, we see that in both cases there is convergence across all library sizes tested. This occurs even though the predictive skill is much higher when NDVI is the causal variable, suggesting a mirage correlation of bimodal causation (Fig. 4).

The results of the Granger causality test from the same analysis can be seen in Table 1. The Granger causality test yielded a few similar results with many notable exceptions. One exception is that, according to the Granger causality test, stream discharge granger causes air temperature and VPD, while CCM does not show these as significant relationships. Also our results show bimodal granger causality with NDVI and discharge, and although CCM shows a similar result, CCM shows the predictive skill of NDVI causing discharge to be twice that of discharge causing NDVI. (Fig. 4, Table 1).

Figures 5, 6, and 7 illustrate the interactive strength of the multivariate EDM. The trends in stream discharge, precipitation, air temperature, vapor pressure deficit, and NDVI are not only shown, but how these trends compare to each other, or the signals between variables, can also be observed in these figures. Figure 5 shows a seasonal variability of the variables across three years of the time series data. Patterns of variable interaction across the seasons of spring,

summer, fall, and winter are obvious. Overall, spring and fall tend to exhibit a greater amount of variability in all variables while summer exhibits the least amount (Fig. 5). Close inspection of these figures yields some interesting observations. During the spring, noticeable trends include that stream discharge is often following the opposite trend of NDVI, and air temperature and precipitation are typically trending in the same direction (Fig. 6). During the summer, there is a noticeable opposite trend between stream discharge and precipitation, and also air temperature and precipitation (Fig. 5). Stream discharge generally increases significantly with sharp decreases in NDVI. Precipitation and temperature tend to trend more heavily in opposite directions during the fall, compared to summer (Fig. 5). During the winter months, stream discharge typically increases with increases in air temperature and precipitation. NDVI however typically decreases with increases in stream discharge, air temperature, and precipitation (Fig. 5).

Of all of the years included in the time series, year 2012 exhibited the greatest variability in trends of interaction strength for predictive variables (Fig. 7). During the spring of 2012, extremes and large fluctuations in all variables were present. Most notable during the year 2012 are that air temperature and precipitation are largely associated with significantly low NDVI and that during the spring of 2012, increases in temperature and precipitation do not consistently increase stream discharge (Fig. 7).

## **Discussion**

Understanding interactions between variables and forecasting the future status of a system are an integral component of ecological studies. Traditional statistical methods, however, often are incapable of this task due to the complex and dynamical nature of ecosystems. The objective of this paper is to evaluate the effectiveness of a recently developed nonparametric approach, EDM, to study complex dynamical systems that may be nonlinear in terms of

relationships between involved variables. For comparison, we also adopted the traditional methods of linear time series analysis to the same dataset. The case study is a hydrologic system of the headwaters of the Tygart river system in northern West Virginia.

Our findings suggest that through the process of convergent cross-mapping, we determined causality between stream discharge, precipitation, air temperature, vapor pressure deficit, and NDVI. The results of convergent cross mapping demonstrate the causal relationships between variables that drive the hydrological cycle of the temperate deciduous forests of northern West Virginia. The incongruent results of the analysis obtained from the Granger causality test substantiate the use of EDM as a supplement to classical time series analysis.

Limitations of time series data length is a challenge that researchers are often faced with. If time series data is not of sufficient length, the intended analysis may not be possible or at minimum, model forecasting ability is reduced. For this reason, variables in many time series datasets are rendered inadequate for use in ecosystem modeling. Our results show that EDM may be more flexible with shorter time series than ARIMA models. Thus, EDM may be a more appropriate means for researchers to model variables that were collected over a short span of time, strengthening their analysis. Additionally, our results illustrate that EDM has higher forecasting power across varying resolution scale than ARIMA models. Thus, the availability of higher resolution time series data may allow for EDM to be used with shorter lengths of time series.

When analyzing multiple time series data, lags of different time series data may be combined to create the high dimensional manifold (Deyle and Sugihara 2011; Sauer et al. 1991). A primary advantage of EDM is to apply it to dynamic systems consisting of noisy observations. Thus, an approach using multivariate time series is likely to be more appropriate than applying a

univariate model when reconstructing the manifold in high dimensional space (Ye 2017). However, it is interesting to note that our results showed that the univariate EDM model had higher forecasting performance than either the multivariate or multiview EDM. Even more intriguing, is to consider how much higher the forecasting ability of univariate EDM is compared to that of the multivariate ARIMA model. This result is best explained by Takens' Theorem which expresses that the entire behavior of a dynamic system is contained in, and may thus be reconstructed from, a single time series variable (Takens 1981). Thus, EDM may be a practical approach to forecasting when only one or two time series variables are available. Furthermore, our results which show that the multivariate and multiview EDM approaches outperformed the multivariate ARIMA model, is in agreement with the concept of multivariate EDM generally outperforming linear models due to the ability of EDM to capture noise in the data and use it to enhance forecasting ability (Ye 2017).

The interactions between variables of time series data are often considered to be important drivers within a dynamic system (Chang et al. 2017). Traditionally, under applications of linear statistical approaches, interaction between variables is typically addressed through implementing impulse response functions, step up, and top down methods and evaluating using AIC, correlations coefficients, and F-values (Horswill et al. 2014; Lütkepohl 1990). However, none of these methods allow for the capture and visual interpretation of time-varying change of the interaction between time series variables (Deyle et al. 2016). EDM allows for this specific analysis to be accomplished through the S-map method in which partial derivatives are calculated for the predictive variables for each time step in a multivariate state space (Chang et al. 2017).

EDM shows the dynamics between variables and how these variables change depending upon the state of the system. In the context of this study, this allows for the identification of shifts in dependent variables as they relate to changes in seasons or climatic conditions. Our results illustrate several key points regarding driving variables, and their interactions, of the hydrological system in northern West Virginia.

High and low air temperatures generally vary considerably during the spring and fall in northern West Virginia as illustrated by our results. During these seasons, as air temperatures increase, there is generally an increase in precipitation due to warm, moist airflow from the Gulf of Mexico. This in turn may cause increases stream discharge. However, if air temperatures remain high enough to enhance greening of vegetation, particularly in the spring, then an increase in NDVI will accompany stream discharge decreases due to an uptake of moisture from the soil from plants.

The summer months generally exhibit the least amount of variability in weather in northern West Virginia. Increases in precipitation during the summer months have less effect on stream discharge due to the abundant green vegetation which consistently removes large quantities of moisture from the soil, putting an upper limit on stream discharge. When summer temperatures are high and humidity levels are low, vapor pressure deficit increases further contributing to an upper limit on stream discharge.

Although winters are generally cold in northern West Virginia, large increases in temperature above average are not uncommon and are often accompanied by above normal amounts of precipitation due to large areas of low pressure crossing the state from the southwest and also stalled frontal boundaries. During these events stream discharge increases significantly. The phenomenon in our results of NDVI decreasing during trends of increasing precipitation and



temperature is likely explained by the concept that such increases in temperature and precipitation are accompanied by snowfall which reduces the amount of greenness, leading to a decreasing NDVI.

The results obtained from our interaction strength analysis are consistent with commonly observed and accepted climatic/weather patterns across northern West Virginia. The fact that accepted climatic conditions and weather patterns are depicted so accurately in our interaction strength analysis illustrates the point that this methodology is very useful to determining unknown driving factors behind unusual trends, or chaotic patterns, in the hydrology of northern West Virginia. For example, during the spring of year 2012, abnormal warmth remained in place over much of the eastern half of the United States (Ault 2013). By analyzing driving variables and their interactions during this period, we can learn about the ecosystem processes and how the hydrology of an eastern deciduous forest may be altered with changing climate and weather patterns.

Understanding the complex and dynamic behavior of an ecosystem is paramount to fully analyzing it. EDM not only allows researchers the ability to understand the driving time series variables and interaction between these variables that comprise ecosystem processes, but also the opportunity to strengthen their forecasting ability by using robust models based on of the complexity and full scope of their datasets. EDM also provides a form of validation to evaluate existing linear models such as ARIMA and the Granger Causality test. Our results illustrate that EDM is an efficient, cutting edge tool for analyzing and forecasting the hydrology of West Virginia's temperate deciduous forest.

## References

- Anderson, C.N., Hsieh, C.-h., Sandin, S.A., Hewitt, R., Hollowed, A., Beddington, J., May, R.M., and Sugihara, G. 2008. Why fishing magnifies fluctuations in fish abundance. *Nature* **452**(7189): 835.
- Chang, C.-W., Ushio, M., and Hsieh, C.-h. 2017. Empirical dynamic modeling for beginners. *Ecological Research* **32**(6): 785-796.
- Daly, C., R.P. Neilson, and D.L. Phillips. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, 33 (2), 140–158, doi:10.1175/1520-0450(1994)033h0140:ASTMFMi2.0.CO;2.
- Deyle, E.R., May, R.M., Munch, S.B., and Sugihara, G. 2016. Tracking and forecasting ecosystem interactions in real time. *Proc. R. Soc. B* **283**(1822): 20152258.
- Deyle, E.R., and Sugihara, G. 2011. Generalized theorems for nonlinear state space reconstruction. *PLoS One* **6**(3): e18295.
- Dickey, D.A., and Fuller, W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root. *journal of the American Statistical Association* **74**(366a): 427-431.
- Glaser, S.M., Fogarty, M.J., Liu, H., Altman, I., Hsieh, C.H., Kaufman, L., MacCall, A.D., Rosenberg, A.A., Ye, H., and Sugihara, G. 2014. Complex dynamics may limit prediction in marine fisheries. *Fish and Fisheries* **15**(4): 616-633.
- Horswill, C., Matthiopoulos, J., Green, J.A., Meredith, M.P., Forcada, J., Peat, H., Preston, M., Trathan, P.N., and Ratcliffe, N. 2014. Survival in macaroni penguins and the relative importance of different drivers: individual traits, predation pressure and environmental variability. *Journal of Animal Ecology* **83**(5): 1057-1067.
- KUNDZEWICZ, Z.W., and NAPIÓRKOWSKI, J.J. 1986. Nonlinear models of dynamic hydrology. *Hydrological sciences journal* **31**(2): 163-185.
- Lütkepohl, H. 1990. Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models. *The review of economics and statistics*: 116-125.
- Nau, R. 2018. Statistical forecasting: notes on regression and time series analysis.
- NOAA. 2018. National Weather Service Forecast Office Charleston, WV. Online Weather Data.
- Pinzon, J.E., and Tucker, C.J. 2014. A non-stationary 1981–2012 AVHRR NDVI3g time series. *Remote Sensing* **6**(8): 6929-6960.
- Sauer, T., Yorke, J.A., and Casdagli, M. 1991. Embedology. *Journal of statistical Physics* **65**(3-4): 579-616.
- Sugihara, G. 1994. Nonlinear forecasting for the classification of natural time series. *Phil. Trans. R. Soc. Lond. A* **348**(1688): 477-495.
- Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., and Munch, S. 2012. Detecting causality in complex ecosystems. *science*: 1227079.
- Sugihara, G., and May, R.M. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**(6268): 734.

Takens, F. 1981. Detecting strange attractors in turbulence. *In* Dynamical systems and turbulence, Warwick 1980. Springer. pp. 366-381.

Ye, H., Adam Clark, Ethan Deyle, George Sugihara. 2017. "rEDM: an R package for Empirical Dynamic Modeling and Convergent Cross Mapping".

Ye, H., Beamish, R.J., Glaser, S.M., Grant, S.C., Hsieh, C.-h., Richards, L.J., Schnute, J.T., and Sugihara, G. 2015. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proceedings of the National Academy of Sciences* **112**(13): E1569-E1576.

Table 1. Summary statistics of all variables.

	Stream Discharge (m <sup>3</sup> /s)	NDVI	Precipitation (mm)	Mean Air Temperature (degrees C)	Vapor Pressure Deficit (hPa)
Mean	370.80	0.65	3.19	10.37	5.31
STDEV	669.41	0.18	6.82	9.06	3.43
Min	0.00	0.24	0.00	-19.20	0.01
Max	15800.00	0.95	93.09	26.20	21.00

Table 2. Statistics of Dickey Fuller Unit Root Test for original data and first differences.

<b>Variable</b>	<b>F-Statistic (Original Data)</b>	<b>F-Statistic (First Difference)</b>
Stream Discharge	-30.7468	-83.8649
Precipitation	-45.7494	-97.468
Mean Air Temperature	-11.5818	-77.1966
Mean Vapor Pressure Deficit	-14.7773	-84.803
NDVI	-4.1131	-30.073

Table 3. Results of Granger causality test between stream discharge, precipitation, air temperature, vapor pressure deficit, and NDVI.

<b>Research Hypothesis</b>	<b>F-Statistic</b>	<b>Significance</b>
Precipitation → Stream Discharge	66.09	***
Stream Discharge → Precipitation	1.40	
Mean Air Temperature → Stream Discharge	7.77	**
Stream Discharge → Mean Air Temperature	278.62	***
Mean Vapor pressure Deficit → Stream Discharge	352.00	**
Stream Discharge → Vapor Pressure Deficit	22.30	**
NDVI → Stream Discharge	109.90	***
Stream Discharge → NDVI	170.96	***



Figure 1. Location of Valley Bend, West Virginia, site of data collection for all variables.

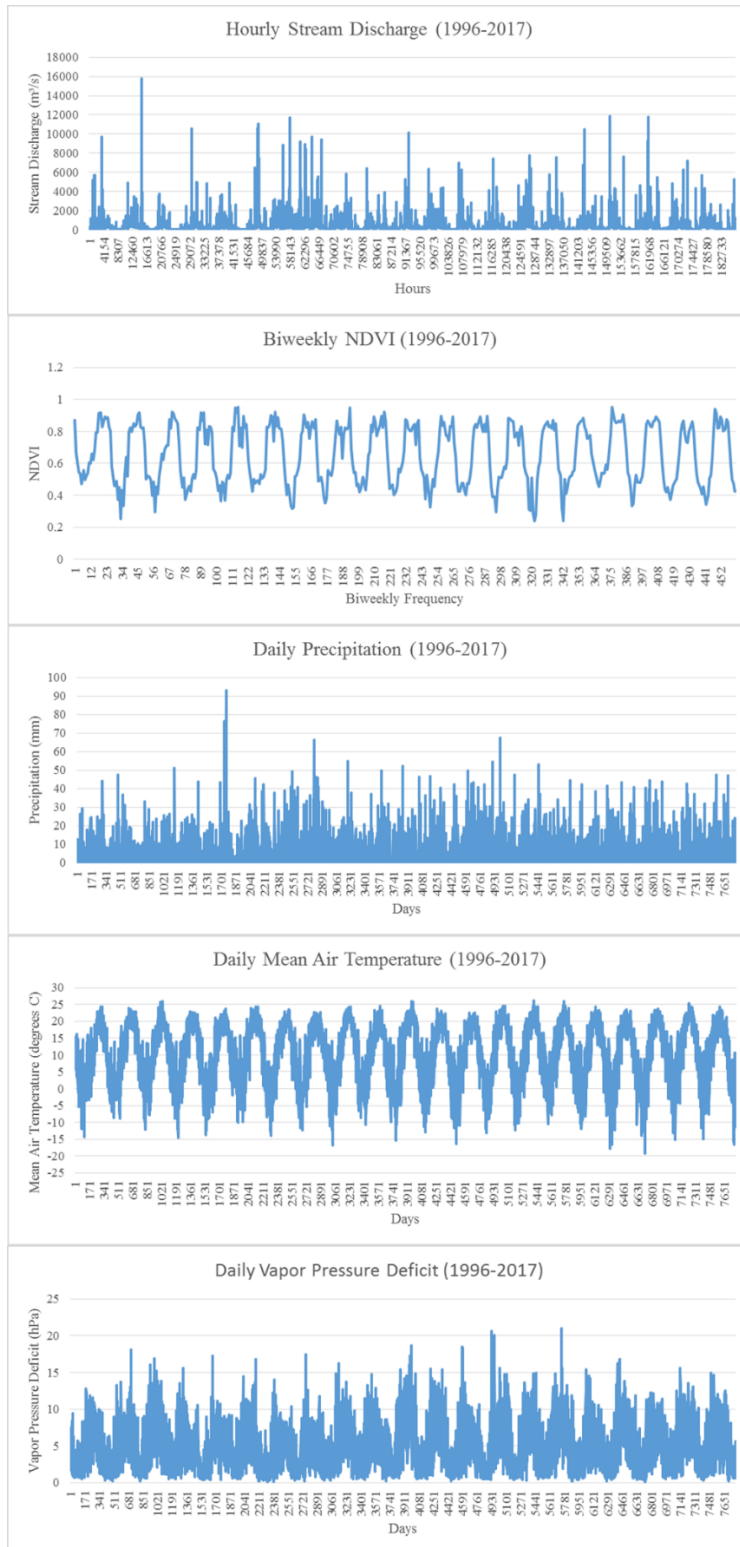


Figure 2. Original stream discharge, NDVI, precipitation, temperature, and vapor pressure deficit data.



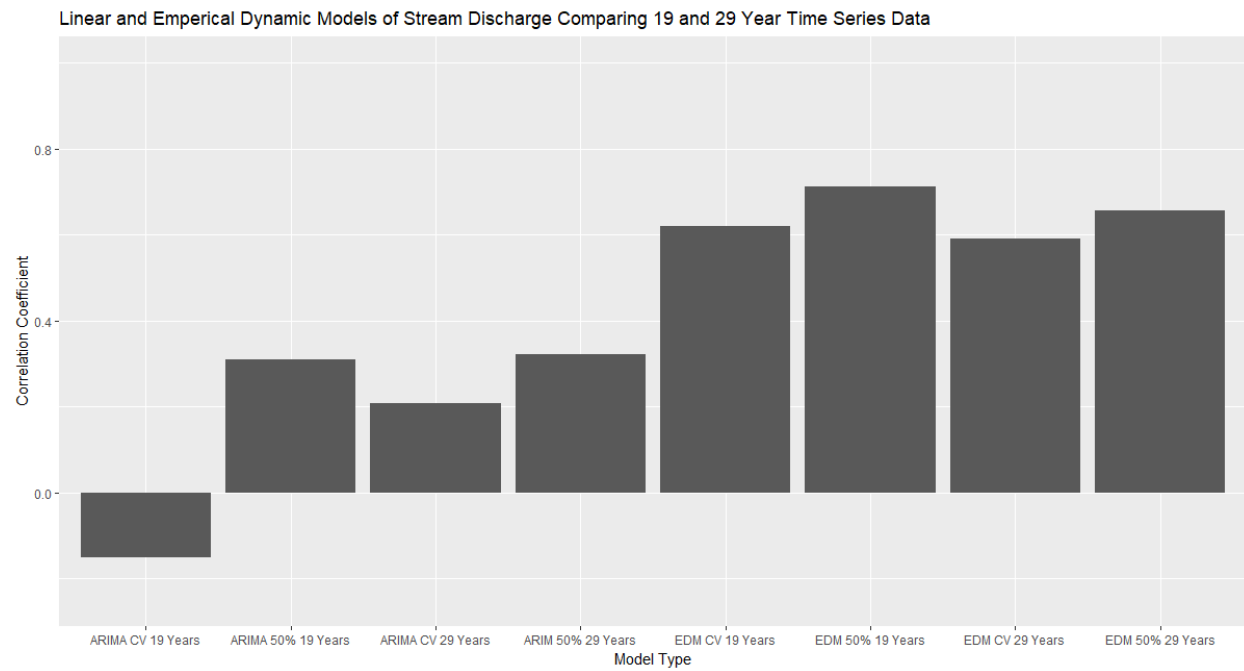


Figure 3. Comparison of time series data length for linear and EDM models of stream discharge at monthly time steps.

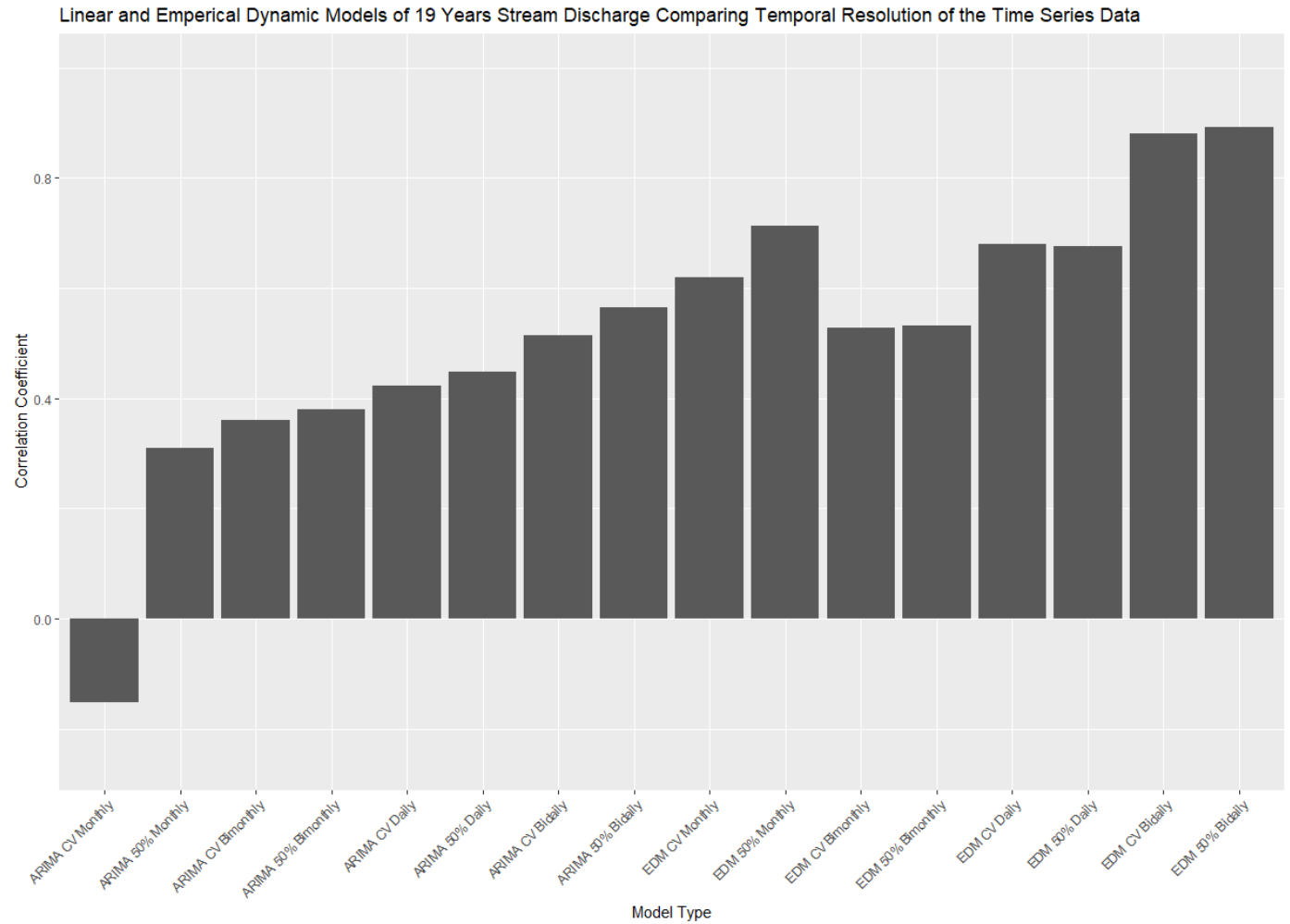


Figure 4. Comparison between resolutions of 19 year length time series data for linear and EDM models of stream discharge.

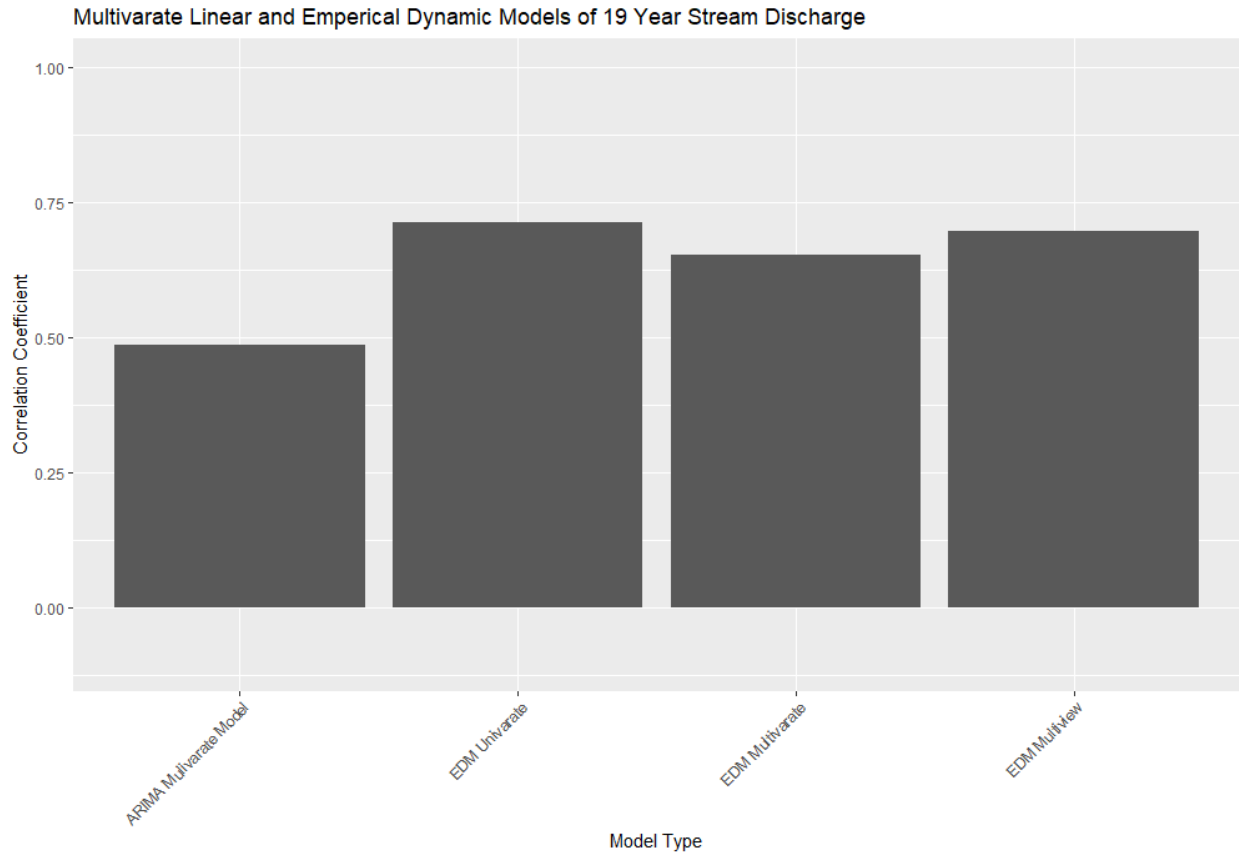


Figure 5. Multivariate linear models and empirical dynamic univariate, multivariate, and multiview models.

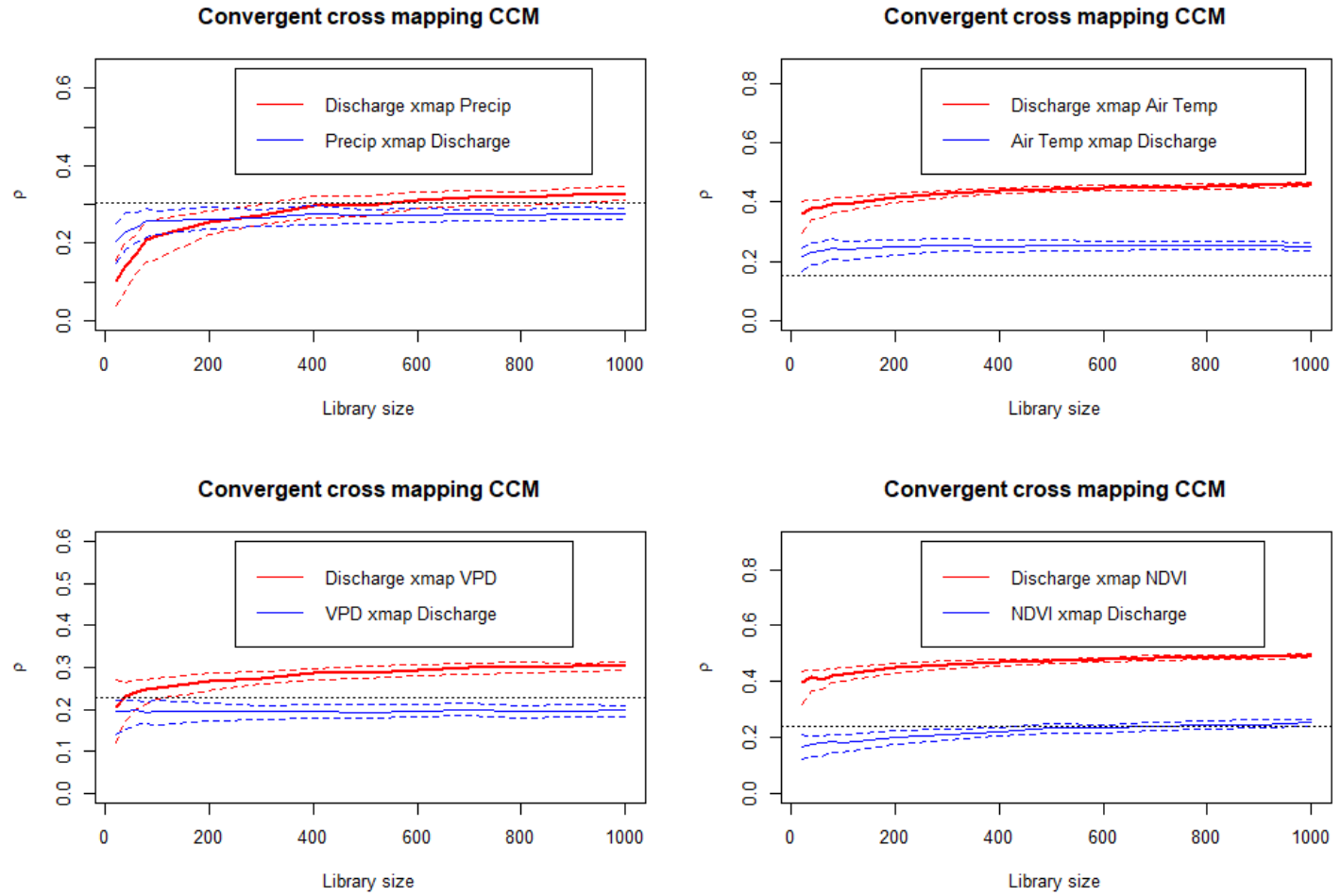


Figure 6. Convergent Cross mapping to identify causality between stream discharge and precipitation, air temperature, vapor pressure deficit, and NDVI.

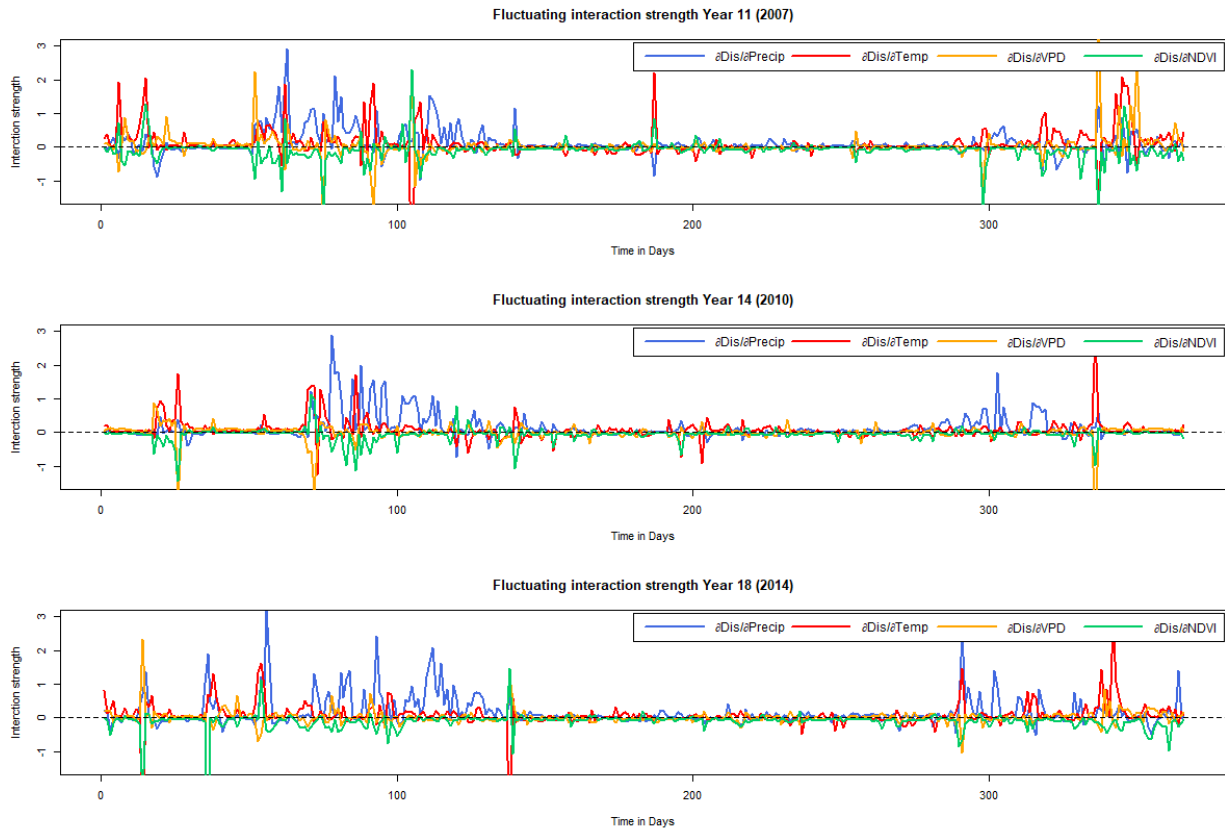


Figure 7. Interaction between stream discharge, precipitation, air temperature, vapor pressure deficit, and NDVI for the years 2007, 2010, and 2014 at Valley Bend, WV.

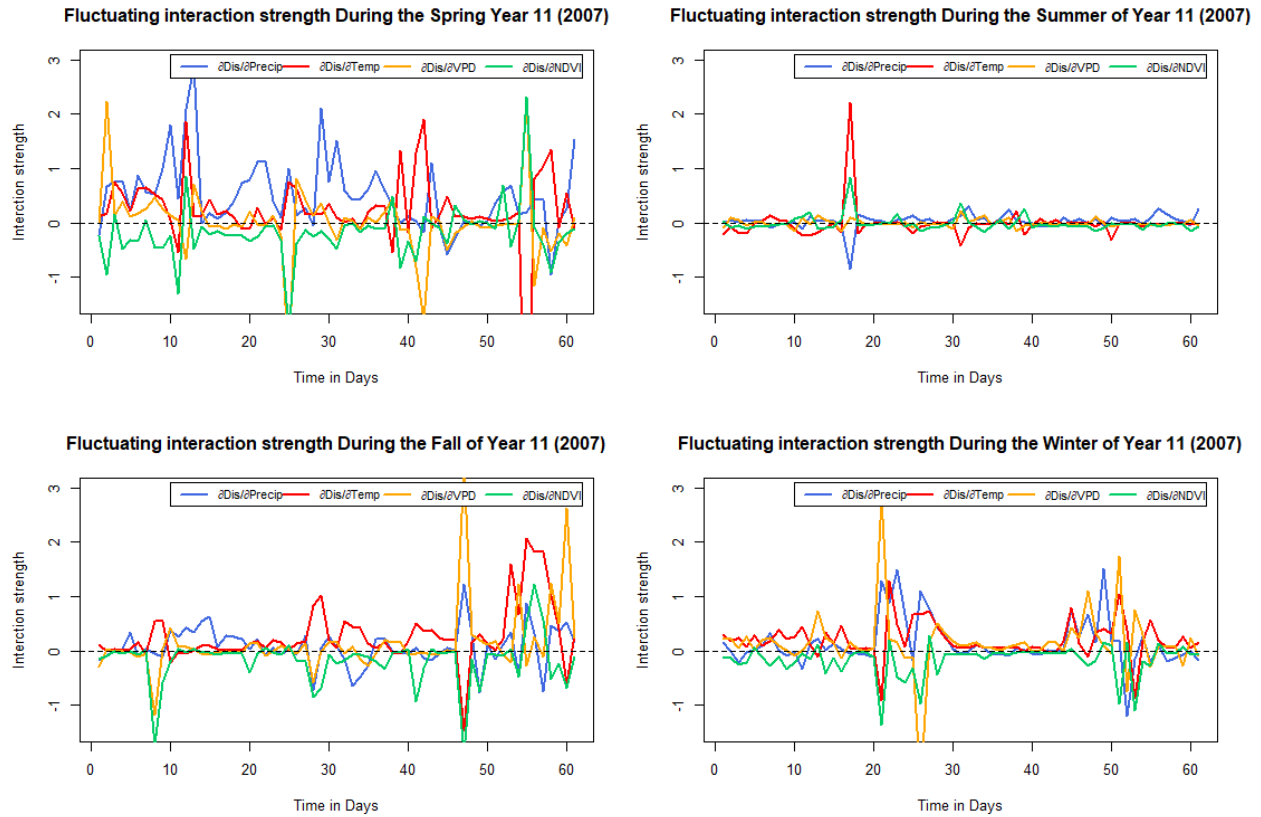


Figure 8. Comparison of seasonal interaction between stream discharge, precipitation, air temperature, vapor pressure deficit, and NDVI for the year 2007 at Valley Bend, WV.

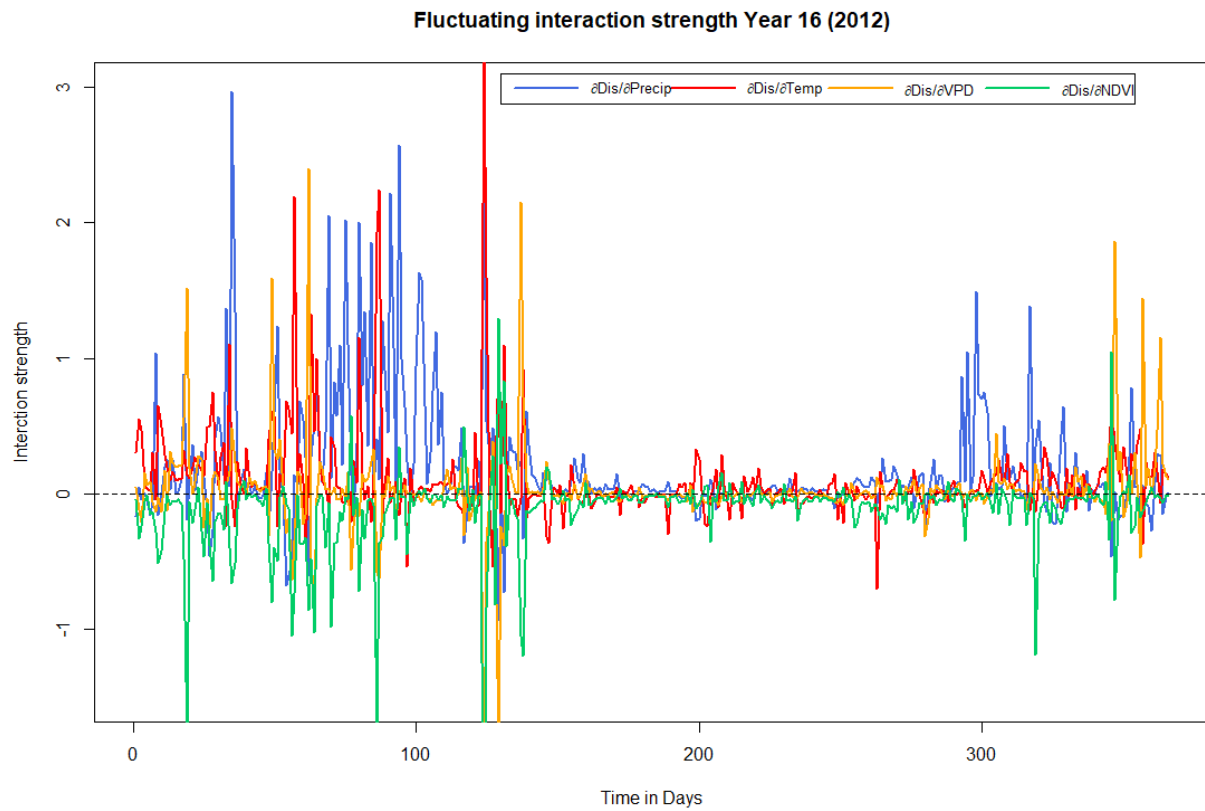


Figure 9. Interaction between stream discharge, precipitation, air temperature, vapor pressure deficit, and NDVI for the year 2012 (a year with significant severe local weather events), illustrating change in these variables under extreme weather patterns.